Winter 12-15-2022

System shock: Using -omics data to characterize biological remodeling in the microbiome, resistome, and transcriptome of bacteria

Winston Eugene Anthony
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Molecular and Cellular Biology

Dissertation Examination Committee:

Gautam Dantas, Chair
Megan Baldridge,
Michael Baym,
Juliane Bubeck-Wardenburg,
Carey-Ann Burnham,
Jim Skeath

System Shock: Using -Omics Data to Characterize Biological Remodeling in the Microbiome, Resistome, and Transcriptome of Bacteria

by

Winston Anthony

A dissertation presented to
Washington University in St. Louis
in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December, 2022
St. Louis, Missouri

Table of Contents

List of Figures viii

List of Tables ix

Acknowledgements x

Abstract of the Dissertation xiii

Chapter 1: 1

Introduction 1

1.1 Defining the healthy gut, its role as a reservoir of antibiotic resistance, and the effect of dysbiosis 1

1.1.1 Abstract 1

1.1.2 Introduction 2

1.1.3 Towards a definition of the healthy gut microbiome 3

1.1.4 The gut as a reservoir of AR 5

1.1.4 Gut microbiome dysbiosis after antimicrobials 7

1.1.1 Techniques for studying the development of AR within the healthy human gut microbiome 9

1.1.9 Conclusion and Future Directions 12

1.1.10 Appendix 13

1.2 Escherichia coli as a chassis for bioproduction of fatty acids 13

1.2.2 Abstract 13

1.2.1 Introduction 14

1.2.3 History of laboratory use of E. coli 15

1.2.4 E. coli as a chassis for bioproduction 16

1.2.5 Conclusion and Future Directions 20

1.3 Development of Rhodococcus opacus as a chassis for lignin valorization and bioproduction of high-value compounds 22

1.3.1 Abstract 22

1.3.2 Introduction 23

1.3.4 Why bio-products? 25

1.3.5 What is lignin? 26

1.3.6 Thermochemical and catalytic conversion of lignin 27

1.3.7 Biological and hybrid conversion of lignin 29

1.3.8 Why *Rhodococcus opacus* PD630 for hybrid conversion of lignin? 31

1.3.9 Genetic and metabolic characteristics of *R. opacus* 33
1.3.10 Tool and technique development for R. opacus engineering 36
1.3.11 *R. opacus* as a production host 43
1.3.112 Conclusion 46
1.4 Acknowledgements 47
Authors' contributions 48
Funding 48
1.4 References 48
Chapter 2: 69
Acute and persistent effects of commonly-used antibiotics on the gut microbiome and resistome in healthy adults 69
2.1 Introduction 69
2.1.1 Abstract 69
2.1.2 Introduction 70
2.2 Results 74
2.2.1 Antibiotics decrease microbiome bacterial load and richness and perturb microbial community structure 75
2.2.2 Azithromycin delays recovery of species richness and is associated with the relative abundance of 9 gut commensals and metabolic pathways 78
2.2.3 The resistance reservoir increases in healthy volunteer microbiomes over time after antibiotic perturbation 80
2.2.4 The healthy volunteer resistome after perturbation is distinct from the resistome of ICU patients 82
2.2.5 Most healthy volunteers remain inside healthy PCA space; some enter ICU PCA space after antibiotics 84
2.2.6 The taxonomy and resistance composition of the healthy volunteer microbiome is altered after recovery 86
2.3 Materials and Methods 88
2.3.1 Resource Availability 88
2.3.2 Experimental Model and Subject Details 88
2.3.3 Semi-quantitative culturing 90
2.3.4 DNA extraction and sequencing 90
2.3.5 Processing of Illumina sequence data 91
2.3.6 Quantification and statistical analysis 91

2.4 Discussion 94
2.4.1 Limitations of the study 99
2.5 Acknowledgements 102
2.5.1 Author Contributions 103
2.5.2 Declaration of Interests 103
2.5.3 Inclusion and Diversity 103
2.6 References 103
2.7 Appendix (Supplemental Material) 111
2.7.1 Figures 111
2.7.2 Tables 114
Chapter 3: 122
Engineering Diverse Fatty Acid Compositions of Phospholipids in Escherichia coli 122
3.1 Introduction 122
3.1.1 Abstract 122
3.1.2 Introduction 123
3.2 Results 126
3.2.1 Engineering E. coli to modulate phospholipid profile 126
3.2.2 Phospholipid profile analysis of the IBFA-producing strain 132
3.2.3 Phenotypic profiling of engineered strains with diverse FA profiles 135
3.2.4 Transcriptomic profiling of engineered strains with diverse FA profiles 137
3.2.5 Engineering *E. coli* for production of free IBFAs 141
3.3 Materials and Methods 144
3.3.1 Reagents 144
3.3.2 Strains and plasmids 144
3.3.3 Fermentation 145
3.3.4 Quantification of total FAs 145
3.3.5 Quantification of membrane lipids 146
3.3.6 High resolution mass spectrometric analysis of IBFA strain cell membrane 147
3.3.6 Quantification of FFAs 147
3.3.7 Metabolic Profiling Assays 148
3.3.8 Time course growth assays 148
3.3.9 Transcriptomics analysis 149

3.4 Discussion 151
3.5 Acknowledgements 155
3.5 References 156
3.6 Appendix (Supplemental Material) 163
3.6.1 Tables 163
3.6.2 Figures 183
Chapter 4: 196
Autologous transcription regulator overexpression for increased triacylglyceride production in *Rhodococcus opacus PD630* 196
4.1 Introduction 196
4.1.1 Abstract 196
4.1.2 Introduction: 197
4.3 Methods: 199
4.3.5 RNA extraction and rRNA depletion: 204
4.3.6 Sequencing Library Preparation and Transcriptomic Analysis: 205
4.2 Results 205
4.2.1 Fatty acid accumulation in *R. opacus* is nitrogen concentration dependent when grown on glucose or phenol 207
4.2.2 Evidence of plasmid loss across multiple strains when grown in phenol and glucose 207
4.2.3 ATR 13 and 20 overexpression induces transcriptional reprogramming in phenol and glucose 207
4.2.4 ATR 13 and 26 increase expression of translation and phenol utilization genes 209
4.2.5 The PaaX-like gene ATR 13 induces upregulation of metabolic pathways related to protein and cofactor synthesis in phenol 212
4.2.6 The PaaX-like gene ATR 13 requires feaR, the phenylethylamine degradation pathway transcriptional regulator, to increase FA titers 214
4.4 Discussion 217
4.5 Acknowledgements 221
4.5 Appendix (Supplementary Material) 222
4.5.1 Tables 222
4.5.1 Figures 223
4.6 References 224
Chapter 5: 228

Adaption to model lignin breakdown compounds results in divergent evolutionary trajectories 228
5.1 Introduction 228
5.1.1 Abstract 228
5.1.2 Introduction 229
5.2 Results 232
5.2.1 Adapted mutants display concentration-dependent aromatic catabolite repression, increased tolerance, and consumption 232
5.2.2 Adaptation changes the transcriptome of *R. opacus PD630* grown in diverse carbon sources 235
5.2.3 MLBP mutants display minimal evidence of a shared transcriptional state after adaption and divergent regulation of the β-ketoadiapate pathway at higher concentrations 237
5.2.4 PVHG6 exhibits higher transcriptomic resilience than mutants grown on less complex MLBPs 241
5.3 Materials and Methods 241
5.3.1 Supernatant compound concentration analysis: 241
5.3.2 Strains and DNA manipulation 242
5.3.2 Growth assay biomass measurement: 242
5.3.3 *R. opacus PD630* culture conditions: 243
3.3.4 RNA extraction and rRNA depletion: 243
5.3.5 Sequencing Library Preparation and Transcriptomic Analysis: 244
5.4 Discussion 244
5.5 Acknowledgements 249
5.5 Appendix (Supplemental Material) 250
5.5.1 Tables 250
5.5.2 Figures 250
5.5 References 253
Chapter 6: 257
Conclusions 257
References 265
Curriculum Vitae 268
Summary/Skills 268
Education 268

Research Experience 268
Research Articles 269
Press, Fellowships, and Awards 270
Presentations 270
Leadership and Student Organizations 271
Mentoring Experience 271

List of Figures

FIGURE 2. 1....76
FIGURE 2. 2....79
FIGURE 2. 3....81
FIGURE 2. 4....83
FIGURE 2. 5....85
FIGURE 2. 6....87
FIGURE 2. 7....101
FIGURE 2. 8S 111
FIGURE 2. 9S 111
FIGURE 2. 10**S** 112
FIGURE 2. 11**S** 114
FIGURE3.1....18
FIGURE 3. 2....33
FIGURE 3. 3....125
FIGURE 3. 4....128
FIGURE 3. 5....132
FIGURE 3. 6....134
FIGURE 3. 7....136
FIGURE 3. 8....140
FIGURE 3. 9....142
FIGURE 3. 10S 184
FIGURE 3. 11S. 185
FIGURE 3. 12S 186
FIGURE 3. 13S 187
FIGURE 3. 14S 188
FIGURE 3. 15S 189
FIGURE 3. 16S 190
FIGURE 3. 17S 191
FIGURE 3. 18S 192
FIGURE 3. 19S 193
FIGURE 3. 20S 194
FIGURE 3. 21S. 195
FIGURE 4.1....206
FIGURE 4. 2**.**....208
FIGURE 4. 3**:** 211
FIGURE 4. 4**:** 212
FIGURE 4. 5**.**....216
FIGURE 4. 6S 223
FIGURE 4. 7**S** 224
FIGURE 5. 1....233
FIGURE 5. 2....235
FIGURE 5. 3....238
FIGURE 5. 4....240
FIGURE 5. 5....250
FIGURE 5. 6S 251
FIGURE 5. 7S 252
FIGURE 5. 8S. 253

List of Tables

TABLE 2. 1S 114
TABLE 2. 2S 115
TABLE 2. 3S 116
TABLE 2. 4S 117
TABLE 2. 5S 118
TABLE 2. 6S 119
TABLE 2. 7S 120
TABLE 2. 8S 120
TABLE 3. 1 138
TABLE 3. 2 163
TABLE 3. 3S 168
TABLE 3. 4S 171
TABLE 3. 5S 172
TABLE 3. 6S 177
TABLE 3. 7S 181
TABLE 4. 1 222
TABLE 5. 1S 250

Acknowledgements

I thank the Dissertation Committee, with special thanks to my advisor Dr. Gautam Dantas, for invaluable guidance provided. I also thank the administrative and technical staff at the Edison Family Center for Genome Sciences & Systems Biology at Washington University in St. Louis School of Medicine, the participants of the Healthy Volunteer stool study, as well as my colleagues and collaborators in the Dantas Lab, the Kwon Lab, the Foston Lab, the Zhang Lab, and the Moon Lab.

I want to thank everyone who is or was a Dantas lab member. Pre-COVID-19, it would be a rare Friday night spent outside the Dantas lab break room. To all the float trips, nights out, and days in the lab. I especially need to thank Aura, Kim, Luke, Erin, Joohee, Rhiannon, Kevin, and Olivia, as they became more than just lab survivors, but friends as well.

I want to thank the members of the WashU Medical School Musical 2018-2020. It is such a unique opportunity at this stage in life to be given a pure creative outlet, and with such a special group of people. I met people I will know for the rest of my life (my wife, more on her later).

I need to thank Dr. Jim Skeath and Cherilynn R. Shadding for their immense support as leaders of IMSD for most of my tenure as a graduate student. Had it not been for their open doors, I would not have finished the program, and I know for a fact that I am not the only student who that holds true for.

It would be remiss of me not to thank the people I met along the way. My IMSD crew, the bar weirdos, Dorota, my dungeons and dragons group, Pat, Sam, and Brit, Jim and Percy. You kept me sane by keeping things a little crazy. For all the nights I don't remember, you are the friends I won't forget.

I cannot end this section without acknowledging the city of St. Louis. You're small (but not too small), cheap, and tons of fun. My first (and last) apartment I rented solo. Walking the labyrinth.

Tower Grove and Forest Park. Living in Shaw. All the bars that no longer exist: Tom's, ITAP (CWE), Culpepper's, Subzero, Dressel's, Llywelyn's, Mike Talaynas JUKEBOX, The Library Annex, Humphrey's, and last but definitely not least, Atomic Cowboy. I gave you money and you gave me alcohol and innumerable good times. And some bad, but you take it.

I thank my mother Luz and father Ethan, Edward my brother, Alexandra my sister, and my cousin Andres. They have given me far more than I will ever be able to repay. I thank my Uncle Oscar, Grandma Hanna and my grandparents Alicia and Alfonso, for being so supportive when I needed them, and the rest of my extended family on both sides. I thank my in-laws, Jim, Carolyn, and my sister in-law Katherine, for bringing me in to your family. I finally thank my wife, Laura, who has kept me sane, fed, and watered. Meeting her and Luna, our little Tiny, was the best thing to come out of my PhD, bar none.

Winston Anthony

Washington University in St. Louis
December 2022

For Carol. You are always near.

ABSTRACT OF THE DISSERTATION

System shock: Using -omics data to characterize biological remodeling in the microbiome, resistome, and transcriptome of bacteria

by

Winston Anthony

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular and Cellular Biology

Washington University in St. Louis, 2022

Professor Juliane Bubeck-Wardenburg, Chair

The human experience is indelibly linked to microbial life. So much so, that the holobiont theory has been coined to define the assemblage of host and microbe as a discrete ecological unit. Perturbation of the commensal microbiome can create niche space for pathogens, which through the concomitant rise of antimicrobial resistance represent an ever-evolving danger. In Chapter 2 of this dissertation, I describe the results of an interventional study designed to directly perturb the healthy gut microbiome and observe the dynamics of taxonomic composition and functional recovery. I observe significant decreases after 5 days of antibiotic treatment in the gold standard metrics used to measure bacterial viability, yet see these same metrics recover to pre-treatment baseline levels within a few weeks. Recovery of species richness does not come without consequences, however, and results in significant functional enrichment of resistance in healthy volunteer microbiomes in three out of four treatment groups. An increase in compositional dissimilarity for taxonomic and resistome composition up to the end of the 6-month study window further confirms the entrenchment of a novel microbiome structure post-

treatment. By looking past standard metrics of microbiome health and diversity, I observe both acute and long-term changes to the taxonomic assembly of commensal bacteria, the resulting consequences on the resistome of healthy volunteers, and identify individuals at greater risk of microbiome dysbiosis after treatment.

Beyond the bacteria residing on and within us are a slew of fascinating organisms which humanity coopts for another reason: bioproduction of chemicals essential for a functioning society. *Escherichia coli* is an excellent chassis for bioproduction of organic compounds due to its fast growth, genetic tractability, and well-understood metabolism. In chapter 3, four *E. coli MG1655 (E.coli)* strains are synthetically engineered to produce altered fatty acid (FA) compositions via the overexpression of novel biosynthesis pathways, resulting in new membrane phospholipid compositions. Two of which are not natively produced by WT *E. coli*. I observe that increased production of cyclopropane FA (CFA) and novel production of internally branched-chain FA (IBFA) results in largely similar growth rates and cell densities as WT. Production of double unsaturated FA (DUFA) results in reduced growth and metabolic output in multiple environmental conditions, as well as a highly perturbed transcriptomic state likely related to an increased need for maintaining iron homeostasis. Overall, I find the *E. coli* chassis tolerates altered or even novel phospholipid compositions while maintaining WT-like growth.

On the opposite side of the spectrum, *Rhodococcus opacus PD630 (R. opacus)* is less genetically tractable, but it's oleaginous nature and incredible metabolic potential have led to efforts to optimize *R. opacus* for degradation of recalcitrant carbon sources. Lignin is an underutilized resource produced from plant matter which *R. opacus* can degrade into the fuel precursor molecule triacylglyceride (TAG). Unfortunately, *R. opacus* only stores carbon as TAGs during nutrient shortage, which limits overall growth and production. In chapter 4 of this

dissertation, we overexpress autologous transcription factors identified using a top-down transcriptome screen and demonstrate increased TAG production when grown in phenol, an aromatic compound commonly found in lignin breakdown products (LBPs). This is directly tied to increased expression of the aromatic catabolism genes of the β-ketoadipate pathway, and expression of the phenylacetic acid (paa) pathway repressor *PaaX*. Using genetic deletion experiments, we demonstrate the existence of a complex functional regulation mechanism for increased TAG production which requires the expression of the *feaR* activator of the phenylethylamine pathway in the +*paaX* background.

Finally, in chapter 5 we use *R. opacus* strains previously adapted to increasingly diverse mixtures of LBPs (MLBPS) to identify adaptive mechanisms for increased tolerance to aromatic compounds. Adapted strains exhibit increased growth rate in MLBPs, and significantly higher utilization of vanillic acid after adaptation. At high concentrations non-permissive to WT growth however, adapted strains exhibit catabolic repression, preferentially utilizing 4-hydroxybenzoate before other carbon sources. Compared to WT grown in a low concentration of MLBPs, adapted strains exhibit little shared differential expression or differential expression of the aromatic degradation clusters and catabolic pathways required for MLBP utilization. It is at high concentrations non-permissive to WT growth, when the effects of adaptation are strongest, that *R. opacus* exhibits divergent DE in the β-ketoadipate pathway. This led to the identification of a putative operon of 8 genes which are similarly divergently DE in all strains, and contain genes likely involved in aromatic catabolism and lipid biosynthesis. Through each chapter of the dissertation, I study the effect of perturbation on microbial systems at the community and cellular level, identifying in each case the emergent properties and mechanisms used for system

resilience, and how this results in recovery, increased bioproduction, and tolerance.

Chapter 1:

Introduction

1.1 Defining the healthy gut, its role as a reservoir of antibiotic resistance, and the effect of dysbiosis

1.1.1 Abstract

The health of the human gut microbiome is often defined in its absence, through the observation of microbiome dysbiosis during illness or other perturbation. This is because the healthy microbiome can exhibit a range of taxonomic and functional compositions while remaining "healthy". One important functional output of the microbiome is antimicrobial resistance (AR), which is intrinsic to some degree in many of the commensal organisms that inhabit the human gut microbiome. Unfortunately, there is a sizeable body of work identifying the healthy gut microbiome as a reservoir for AR. While there is little danger from AR commensal bacteria, colonization of the microbiome by pathogens during dysbiosis can lead to horizontal transmission and spread of AR, using humans as vectors. Here we define the characteristics of a healthy microbiome and describe the β-lactam and plasmid-mediated quinolone AR reservoir found within the healthy gut. From there, I discuss the concept of gut microbiome dysbiosis and its implications for antimicrobial exposure, and colonization of *Clostridiodes difficile* and multidrug resistant *Enterococcus*. Finally, I summarize the current

techniques for identifying dysbiosis and AR within the gut microbiome and introduce future directions for microbiome research.

1.1.2 Introduction

The definition of the healthy microbiome is, necessarily, quite vague. This is primarily because there are so many caveats involved when attempting to boil down the vast range of microbial compositions seen in the gut microbiome of humans. Is a healthy microbiome one without pathogenic bacteria such as Multidrug resistant *Enterococcus* (MDRE) or *Clostridiodes difficile,* each known to cause infection and death(1, 2)? The answer to this question becomes increasingly muddy when gut microbiomes can exhibit asymptomatic carriage of *C. difficile*(3-5), and considering recent research identifying strains of other known pathobionts such as *H. Pylori* as being protective against esophageal adenocarcinoma(6). Moreover, the "average" composition of the microbiome differs greatly due to changes in environment(7, 8) and diet(9, 10). All of this combines to make interpreting the effects of antimicrobials on the gut microbiome a challenge; any proposed model has the herculean task of integrating specific, testable mechanistic rules alongside broad, repeatedly observable effects on composition and diversity.

Antimicrobials have transformed the practice of medicine, but from the moment of their discovery, their effectiveness has been compromised by the emergence of antimicrobial resistance (AR)(11). AR can be encoded for on antibiotic resistance genes (ARG), or antibiotic target mutations. These mutations may be intrinsic; either disseminated through microbial communities via vertical inheritance, or horizontally via mobile genetic elements (MGE) and extrachromosomal plasmids. Historically, AR was predominately described in pathogens isolated from people with clinically significant infection. It is now known that AR can reside in

organisms isolated from the microbiomes of asymptomatic people, and can later contribute to infection when specific conditions create a permissive niche for the organism to contribute to disease (12).

The gut is a prime reservoir for AR organisms(13). When healthy, the gut microbiome is a stable, diverse community which provides important benefits to the host such as nutrient acquisition and protection from pathogens(14). Antibiotics can perturb this ecosystem by changing its taxonomic and functional composition, creating opportunities for pathogen colonization (15). This "dysbiosis" can allow for AR colonization, increased ARG burden, and enable subsequent AR pathogen invasion into the blood stream, urinary tract, and other organ systems(16). Thus, it is becoming increasingly important to understand how dysbiosis can drive AR in the gut microbiome, and how to prevent or reverse dysbiosis.

Here I first define the universal characteristics of a healthy gut microbiome, discussing both the stability as well as the compositional heterogeneity of the microbial community. From there, the reservoir of β-lactam and plasmid-mediated quinolone resistance in the healthy human gut is presented. The concept of microbiome dysbiosis is examined and I discuss how disruption leads to eventual colonization by the pathogenic organisms *Closteridiodes difficile* and multidrug resistant *Enterococcus.* Finally, the standard methods and technologies for detecting and characterizing antimicrobial resistance genes within the gut microbiome are described, along with the potential for future microbiome-directed methods to detect and prevent infection.

1.1.3 Towards a definition of the healthy gut microbiome

The healthy gut microbiome is a complex, diverse community which is resistant to colonization and proliferation by pathogens(17). When in equilibrium, each taxa effects, and is

affected by, the environment and other taxa round them, promoting a stable ecosystem(18). These interactions between microbial taxa and the environment (I include the host as an aspect of the environment), can take many forms: interspecific interactions such as cooperation(19), competition for resources(20, 21), or predation(22) and abiotic interactions between taxa and the local environment(23). A "healthy" microbiome is thus the emergent property which results from the total sum of these interactions; as such the gut microbiome becomes a continually propagating ecosystem within the host, from which the host can benefit(20)(see (14) for the related concept of the microbiome as ball rolling on a stability landscape).

One level of abstraction beyond niche organization lies the enterotype(24, 25). As the large-scale analysis of the microbiome became cost-effective(26), the concept of the enterotype was an *ad hoc* attempt to reconcile a universal definition of a healthy microbiome with the broad scale of variation observed in taxonomic composition(24). Arumugam et al. 2011 analyzed the composition of microbiome samples across different populations and identified three distinct population structures identified by the most prevalent bacterial genera: 1, *Bacteriodes*, 2, *Prevotella*, and 3. *Rhuminococcus(24).* Though there is considerable compositional variation empirically even when categorizing microbiomes by enterotype, these keystone species further refine the definition of a healthy gut microbiome. Repeated analyses not only confirm their validity, but also their stability(27, 28).

Within this characterization of the healthy gut microbiome as a stable unit of ecology there is a considerable amount of temporal flux, as the gut microbiome is constantly adjusting to external input or stimuli(23). Unlike a closed environment(29), the gut microbiome is exposed to repeated pulses of nutrients and chemicals via digestion of host-consumed food. Compositional difference in the microbiome has been observed due to age(30, 31), time(23, 32) and diet(28).

While the effect of this is quantifiable and can significantly alter the metabolic output of the system, the healthy gut microbiome exhibits resilience and is able to minimize the effect of small perturbations to its composition (33). These alterations pale in comparison to the effects of more harmful stimulus such as antimicrobial induced dysbiosis(28, 34, 35).

1.1.4 The gut as a reservoir of AR

Gut commensal organisms have been previously thought to be innocuous. Breakthroughs in sequencing technology and techniques for determining function and transfer capability are revealing a more nuanced picture of the role of commensals in the gut resistome(36). Even more concerning is the fact that mobile elements such as plasmids can be readily shared between commensal and pathogenic species(37). Here we will discuss β-lactam and plasmid-mediated quinolone resistance, because of their propensity to be located on MGEs facilitating their spread, the ubiquity of β-lactam use around the world, and the importance of β-lactam's as essential treatment options for many different types of bacterial infections.

B-Lactam Resistance in the gut microbiome

β-Lactams are the most commonly prescribed antibiotic class worldwide(38). Microbiomes from 30,000 year old permafrost revealed enzymes within the TEM family, which confer resistance to β-lactams(39). Many of the commensal organisms of the human gut microbiome are carriers of β-lactam resistance genes(40): In one study, 11 strains of *Bacteriodes distasonis* and *Bacteroides vulgatus* carried the β-lactamase *cfxA*. Sequence similarity to *cfxA* genes from other members of the *Bacteriodes* genus suggests widespread dissemination of the gene(41). TEM β-lactamases are a family of enzymes which are often located on plasmids and confer resistance to early cephalosporins and penicillins (42). β-lactamases are thus easily spread; there is evidence of β-lactamase transmission through human vector transmission

networks. In one study 12/18 Swedish students tested negative for ESBL-producing bacterial isolates in the gut microbiome before travel, but later tested positive for ESBLs after travel to India(43).

Widespread range and transmission of ESBLs via plasmids has been identified, with community-associated ESBL infections in the US accounting for over 1/3 of total ESBL infections (44). *Faecalibacterium prausnitzii* and *Prevotella copri* isolated from fecal samples of healthy adults was found to be resistant to the cephalosporins ceftriaxone and cefotaxime(45). Metagenomic analysis of the sequenced isolates found that many of their AR genes were located near mobilization elements such as integrases or on plasmids, indicating evidence of gene transfer.

The gut as a reservoir of plasmid mediated quinolone resistance (PMQR)

The primary method of resistance to quinolones arises in the form of single nucleotide polymorphisms (SNPs) located in areas termed quinolone resistance-determining region, but the last few decades have revealed a new method of quinolone resistance: PMQR(46). Travel to an area of high endemic resistance can act as a vector for transmitting PMQRs. Travelers from The Netherlands had significant acquisition of PMQRs after returning from Southeast Asia and India(47). Phylogenetic studies of this family of enzymes confirmed that PMQRs can be found in soil microbiomes and the gut microbiomes of chickens and humans, suggesting an ecological niche to which it is endogenous(48).

It should be noted that there is an important distinction between species which have intrinsic versus acquired resistance(49). Many important Gram-positive gut commensals are intrinsically resistant to quinolones, and acquisition of quinolone resistance can occur in

commensal *E. coli* after antimicrobial exposures (50). Recent work has elucidated more about the origins of quinolone resistance in the gut microbiome. The chromosomal ancestral source of *qnrB* is theorized to be *Citrobacter*; 37 *Citrobacter freundii* isolates from a Massachusetts hospital contained only *qnrB*, with only two showing the ability to transmit this resistance through conjugation(51). There are several *Citrobacter* commensals in the gut, suggesting that it may be an endogenous reservoir for low level quinolone resistance. There remains much to learn about the range of the AR reservoir in the microbiome, and its origins.

1.1.4 Gut microbiome dysbiosis after antimicrobials

When a healthy gut microbiome is perturbed so greatly that it cannot return to a similar baseline state, it can enter a new, *dysbiotic* state(14). The dysbiotic state is defined by characteristics pre-disposing the host to poor health outcomes(34, 35). Antimicrobials, though necessary for resolving infection, are some of the most effective contributors to microbiome dysbiosis(35, 52, 53). This state is generally characterized by lowered diversity, altered functional output (often implements as increased antimicrobial resistance), and increased susceptibility to infection with pathogens or disease. Hindering a coherent understanding of the true effects of antimicrobials on the gut microbiome are frequent comorbidities in observed patients. Research on the effects of antimicrobials on the heathy gut microbiome has been conducted but is hampered by low n and the use of dated techniques(52, 53).

Colonization resistance prevents invasion by pathogenic bacteria

It is thought that "colonization resistance" occurs either through direct competitive interactions between bacteria, or indirectly through commensal bacteria triggering a host response against pathogens (54). Antimicrobials have been shown to cause disruptions to the gut microbiome by lowering the bacterial diversity of the gut microbiome and thereby allowing

pathogens to invade(15). Once this diversity has been compromised it can be difficult to ameliorate(15).

C. difficile colonization can lead to symptomatic C. difficile infection (CDI)

C. difficile is an organism that can cause asymptomatic gut colonization, but can also proliferate in the presence of antimicrobials, leading to CDI. Asymptomatic carriage of *C. difficile* without evidence of any clinical infection occurs in 4-15% of adults(55, 56). These asymptomatic carriers may be at higher risk for CDI, and those with an initial episode of CDI are at high risk for recurrent CDI (55).

As confirmed by large observational studies and mouse models, antimicrobial-induced dysbiosis of the gut microbiome is the greatest risk factor for *C. difficile* colonization and CDI(57). In a mouse model, a single dose of clindamycin reduced the diversity of the gut microbiome for 28 days, enabling susceptibility to CDI for up to 10 days(15). It is thought that the structural changes caused by CDI impact the metabolic output of the microbiome(58), and loss of nutrient competitors may decrease colonization resistance(59). These results are concordant with metagenomic research in human cohorts, which have described the human CDI microbiome as lower in diversity, reduction in *Bacteroides* abundance, and increases in *Proteobacteria(60).*

Gut colonization by MDRE can lead to clinically significant infections

The *Enterobacteriaceae* consist of a large family of Gram-negative organisms which can colonize the gut microbiome and cause infections(61). Multidrug resistance in these organisms have been well-described, including extended-spectrum β-lactamases (ESBL) (62). In a Finnish study, 18% of *E. coli* isolated from the stool of asymptomatic, healthy people were found to be

multidrug resistant(63). There is evidence that MDRE risk can be influenced by environment factors: In another study of a healthy immigrant population, it was noted that 20.4% were positive for MDRE upon arrival, and 9.4% were positive after a median of 35 months post-immigration to a country which had significantly less prevalence of endemic resistance(64, 65).

MDREs can colonize the gut and translocate to other body sites and cause clinically significant infections(66). In a prospective study of patients with urinary tract infections (UTI) longitudinal stool and urine samples were collected from patients and cultured for MDREs(67). Clonal analysis of isolates demonstrated there were multiple transmissions of *E. coli*, *Proteus mirabilis*, and *K. pneumoniae* between the gut microbiome and the urinary tract, and that prior to a clinically significant UTI, there was an increased concentration of a clonal organism in the gut.

1.1.1 Techniques for studying the development of AR within the healthy human gut microbiome

Metagenomic analysis of the gut microbiome is rapidly expanding our knowledge of AR, uncovering an incredible diversity of AR genes and plasmids which can be transferred to other organisms within the gut(68). Rapidly expanding efforts to develop microbiome directed diagnostics and therapeutics are creating a need to characterize and quantify the role of the gut as a reservoir for ARG carriage and exchange. Functional metagenomics is a culture-independent approach to uniquely characterize both known and uncharacterized ARGs(69). Functional metagenomics and can serve as both a discovery engine for cryptic and emerging AR, as well as to model the risk of horizontal gene transfer of AR(70). The recent expansion of long-read sequencing (LRS) technologies offers a powerful complement to functional metagenomics, as it can associate ARGs with their host bacteria and mobilization elements, enabling accurate

estimations of how ARs are exchanged between bacteria (71). In concert with microbiologic culture, these culture-independent technologies hold the potential to improve our understanding of AR.

The gut is host to many bacterial species which are difficult to culture via traditional microbiologic methods, making it difficult to investigate their contribution to the AR reservoir(36). Functional metagenomics is a high throughput culture independent approach to assay the functional activities of microbial communities, enabling the functional genetic surveillance of difficult to culture organisms(36, 72). In functional metagenomics, the total microbial community DNA is transformed into a culturable indicator strain (e.g., *E. coli*) which is then phenotypically screened for acquired resistance to different classes of AR. Through this method, functional metagenomics can model mobilizable AR risk by estimating the resistance elements that can be functionally utilized by an organism such as *E. coli*. For each organism of interest, large amounts of genetic material can be simultaneously assayed, and acquired phenotypic resistance profiles generated. Importantly, this technique does not rely on novel AR genes sharing sequence identity to known AR determinants.

Functional metagenomic analysis suggests that AR genes in pathogens are more frequently co-localized with mobility elements than AR genes in environmental microbiomes(73). A research group functionally validated over 1000 AR genes from fecal and environmental microbiomes using this technique, over 10% of which were novel(74). A recent report interrogating the resistome in wild and captive gorillas, chimpanzees, and co-localized humans found AR genes near MGEs with high sequence similarity from all three sources(75). These data suggest that the microbiomes of wild and captive animals may be important reservoirs of AR. Functionally validating AR genes unlocks a better understanding of the AR

reservoir but does not identify the original bacterial host. Thus, a complementary method is needed to characterize the broader genomic context of AR in the microbiome and identify the greatest clinical threats.

Surveys of ARs in the gut microbiome have primarily been accomplished with "short-read" sequencing (SGS) after shearing total DNA into small fragments. These short reads (<500 base pairs) are generally insufficient to assemble circular contigs which can distinguish between chromosomal and plasmid DNA, though recently developed technologies can identify integration of other types of MGE(76, 77). In contrast, long read sequencing (LRS) can generate reads of 10s of kb in length, which can resolve repetitive regions and generate high quality reference assemblies (78). In one study, Bertrand *et al.* applied a hybrid sequencing approach using SGS and LRS to gut microbiome samples, enabling assembly of species genomes from the metagenomes of patients who underwent antibiotic therapy(71). They discovered multiple plasmids unknown to the medical community, and new regions of multi-drug resistance within bacterial species; among these were multiple combinations of carbapenemases co-occuring with ESBLs(79). One new region conferred resistance to carbapenems, aminoglycosides, trimethoprim, and sulfonamides. Previously, this region was not able to be assembled by SGS due to repeat regions, highlighting the opportunity LRS provides to investigate the AR reservoir of the gut microbiome.

LRS is creating new opportunities to investigate understudied vectors of AR. A recent work identified two new megaplasmids (>420kb) carried by *Pseudomonas aeruginosa* clinical isolates harboring a shared core genome and varying AR gene carriage. GenBank homology searches revealed 72 more bacteria harboring similar megaplasmids, isolated from all over the world, and as far back as 1970. New methods are currently being generated for determining the

genomic context of AR in the gut using LRS, and it is even being integrated into established metagenomic pipelines, increasing the ease-of-use and potential incorporation into clinical practice(80, 81).

Utilizing both functional metagenomics and LTS can reveal nuanced and even more interpretable relationships between AR and the microbiome. In a remarkable study, Kintses *et al.* used functional metagenomics to describe the reservoir of antimicrobial peptide and AR genes, then used LGS to contextualize genes to mobile elements(82). Their investigation revealed that phenotypic resistance in *E. coli* via AR is much more likely to be successfully transferred, located on mobilizable genetic elements, and has fewer phylogenetic barriers to transfer. This is an interesting finding given that the gut microbiota is a known reservoir of antimicrobial peptide genes, and their prevalence was similar to AR genes after selection(83). Both LGS and functional metagenomics, and especially a synergy of the two techniques, enable unparalleled insight into context and function of the AR reservoir. They are an invaluable resource as we move towards a future of microbiome-directed methods of identifying and preventing the spread of AR and infection.

1.1.9 Conclusion and Future Directions

The pace of AR is quickly overtaking the discovery of new antimicrobials, thus continued research into the gut microbiome as an AR reservoir is also of utmost importance. Key areas for future investigations include clinical and translational studies delineating the features in the gut microbiome that are permissive to, or protective against, gut colonization with AR organisms. To achieve this, long-term follow up of asymptomatically colonized people is needed to assess the risk for increased AR after treatment with antimicrobials. Identifying risk factors for increased

AR carriage after certain antimicrobials and could lead to direct improvements in patient care via personalized medicine.

1.1.10 Appendix

This work was produced in collaboration with Carey-Ann D Burnham, Gautam Dantas, and Jennie H Kwon and is adapted from the manuscript published in the *Journal of Infectious Disease*(84). We thank everyone in the Burnham, Dantas, and Kwon labs for their helpful discussions relating to this work.

Disclaimer. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Financial support. This work is supported by the National Institute of Allergy and Infectious Diseases, National Institutes of Health (grant 1K23AI137321 to J. H. K.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Supplement sponsorship. This work is part of a supplement sponsored by the National Institute of Allergy and Infectious Diseases (NIAID), part of the National Institutes of Health (NIH) and the Centers for Disease Control and Prevention (CDC).

1.2 Escherichia coli as a chassis for bioproduction of fatty acids

1.2.2 Abstract

Escherichia coli is a model organism widely used as a vector for cloning and expression, basic research, as well as for bioproduction. Due in part to its early discovery and ease of growth,

E. coli has cemented itself as the bacteria of choice for many of the scientists who went on to author paradigm-shifting research. The discovery and sequencing of the genome has supercharged the use of *E. coli* as a chassis for bioproduction. One area of bioproduction in which *E. coli* holds enormous promise is in the bioproduction of fatty acids and other value added products. In this section, I describe the history behind *E. coli*'s adoption within the scientific community. Focusing on recombinant protein production and bioproduction of phospholipids, I detail the advances made by the scientific community using the bacterium to provide society with medicines and commodity chemicals. Finally, I conclude this section of my dissertation with the current challenges we face adapting *E. coli* for increased bioproduction of phospholipids.

1.2.1 Introduction

Ask anyone who has spent time working in an experimental biomedical laboratory, and you would be hard-pressed to find someone who has never grown, cloned, or sequenced *E. coli*. No matter the department or area of research, *E. coli* reaches beyond the concept of a model organism to a tool for basic biological science.

Scientists have been isolating and studying this enteric bacterium as far back as the 1880's(85), and some of the most fundamental breakthroughs in molecular biology(86), biochemistry(87), and genetics(88) were elucidated through the lens of this hardy, fast-growing, and easy to study organism. It is because of *E. coli*'s genetic tractability and fast growth that it is considered a promising chassis for the bioproduction of phospholipids, recombinant proteins, and biofuel precursors(89-91). Looking forward, *E. coli* as a cell factory holds much promise as we move into the upcoming decades with a focus on sustainable and cost-effective biofuel production(92). In this section, I give a brief overview of the history of *E. coli* as a model

organism, its place in scientific history, and the attributes that make it so desirable as a vector for research. Then I describe the current efforts in the field to modify the *E. coli* chassis for bioproduction. Finally, I focus specifically on *E. coli* for bioproduction of lipids and fuel precursors, and on the challenges which still need to be addressed.

1.2.3 History of laboratory use of E. coli

The love story between biological research and *E. coli* begins in 1884, when the German microbiologist Theodore Escherich began characterizing a bacteria isolated from the intestines of neonates(85). Research using the then-named *Bacterium coli* eventually moved from descriptive ecological studies(93, 94) into the laboratory of 20th century microbiologists(95). *Bacterium coli* demonstrated fast growth and ease of cultivation, enabling breakthrough research on vitamins(87), and bacteriophage(96). Eventually renamed *Escherichia coli*, it was used by Francis Crick to demonstrate the three letter "code" of the genome(88, 97), and elucidate the nature of RNA(86) and restriction enzymes(98). Moving forward into the era of modern genetics, the creation of the first functional bacterial plasmids occurred in *E. coli*(99) and in 1997 the first genome of *E. coli* was sequenced, leading to an explosion of genetic analysis using the bacteria(100-102). The RefSeq(103) database currently lists over 31,000 assembled *E. coli* genomes and 6500 plasmids, with a median genome length of 5.11174 Mb. On paper, *E. coli* is about as close to ideal a model organism as one can get. It has a doubling time of 15-20 minutes at 37C, and the complete metabolic pathway databases(104) for *E. coli* are compiled for anyone in the world to use. The *E. coli* K12 proteome(105), metabolome(106), and regulome(107) are also well described. All of this results in a staggering amount of data: a pubmed search for "*Escherichia coli*" gathers nearly half a million related publications and positions the bacteria well for use in commercial bioproduction applications.

1.2.4 E. coli as a chassis for bioproduction

Bioproduction of recombinant protein therapies in *E. coli*

Even before a full genome sequence of *E. coli* was published scientists were using recombinant DNA to add novel genetic information. This eventually culminated in the production of human insulin from recombinant DNA(108). Inevitably federal guidelines for the responsible use and development of products involving recombinant DNA in *E. coli* K12 were established (and are periodically updated)(29). This has allowed for production of a slew of *in vivo* synthesized therapeutic proteins to be brought to market treating gout, Crohn's disease, and multiple sclerosis(109). These are only a sampling: As of 2012, a third of all recombinant therapies are produced in *E. coli(109)*.

There have been further modifications to the *E. coli* chassis beyond what is required for target production to reduce off-target consequences of large-scale protein production and improve efficiency. The toxic anion acetate forms as a byproduct of aerobic respiration and reduces protein production when present. Deletions of the phosphotransferase system (PTS) genes reduce the uptake of glucose, resulting in a elimination of acetate formation and a significant reduction of growth rate, but was successful in increasing overall recombinant protein production(110). This is indicative of a fundamental trade off in bioproduction: To some degree the negative consequences of genetic modifications can be considered tolerable if offset by production gains. Another method of acetate reduction is the introduction of biosynthetic pathways which divert glycolysis products such as pyruvate away to acetoin, resulting in a 60% increase in recombinant protein activity(111). Introduction of pyruvate carboxylase from *Rhizobium etli*, an oxaloacetate synthesis protein, successfully decreased acetate in the media by 43%. More research is needed to fine tune genetic modulation of acetate production. One method

of interest is using bulk expression profiling techniques to observe the resulting changes to all metabolic pathways of the cell(112, 113). Secondary modifications can then be attempted from using insight from the RNA expression data.

Bioproduction of fatty acids and biofuel precursor molecules in *E. coli*

Phospholipids are essential components of the bacterial cell, comprising the semipermeable membrane which encloses the cellular compartment(114). The bacterial cell membrane plays important roles in nutrient transport as well as energy production though the generation of a proton motive force, and thus phospholipid content can modulate many aspects of microbial cell health(115, 116). Apart from their roles in biology, phospholipids are highly prized by humanity as well: they are the precursor molecules for detergents, commodity chemicals, and biofuels. Phospholipids are composed of a hydrophilic head group containing phosphoglycerol and a hydrophobic fatty acid acyl chain (see (117) for a review of the synthesis of phospholipids).

The phospholipid content of Gram-negative bacteria is high in saturated fatty acids, with smaller contingents of unsaturated fatty acids and cyclopropane fatty acids(118). Fatty acids vary in structure, which in turn imparts different cell membrane characteristics, and the cell is constantly modulating FA membrane composition. *E. coli* modulates the specific quantities of each of the three fatty acids (FAs) in response to stress and other environmental variables, such as by increasing the proportion of cyclopropane fatty acids during acid or cold stress(119, 120). Purpose-driven microbial adaption and engineering to change characteristics of bacterial cells is a concept called “membrane engineering” and is being used to modify *E. coli* for bioproduction of fatty acids(120-122). The *E. coli* type II fatty acid biosynthesis genes have been extensively characterized(92), and fatty acid biosynthesis begins with the production of malonyl-CoA by the

accABCD acetyl-CoA carboxylase complex. From there the malonyl moiety is transferred to ACP via *fabD,* is then condensed into acetoacetyl-ACP by *fabH,* and finally enters the fatty acid elongation cycle. Both *fabH* and *fabD* are essential to fatty acid biosynthesis, and the latter to *E. coli* growth(123, 124). *fabA and fabB* are then responsible for producing unsaturated fatty acids and are regulated by the fatty acid regulators *fadR* and *fabR*(125, 126). Further rounds of double bond and methyl groups modifications to produce other forms of fatty acids such as occur after phospholipid synthesis and translocation to the membrane(127, 128).

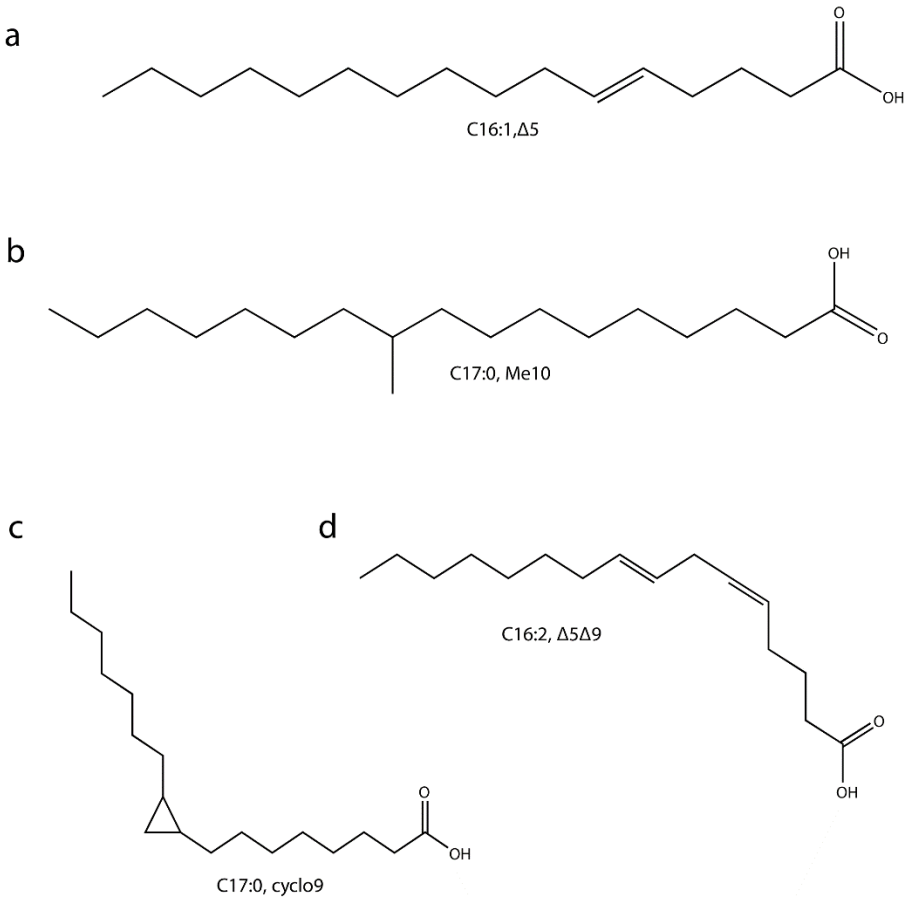


*Figure 3. 1****Representative fatty acids from bacteria. a. Mono-unsaturated fatty acid (UFA), b. branched-chain fatty acid (BCFA), c. cyclopropane fatty acid (CFA), d. double unsaturated fatty acid (DUFA).***

Bioproduction of PUFAs in *E. coli*

Poly-unsaturated FAs (PUFAs) are highly prized due to possessing commercially desirable characteristics over mono-unsaturated FAs(129). PUFAs scavenge radical oxygen species and are marketed as health supplements, but commercial supply primarily relies on phytoplankton-fed fisheries which are increasingly unable to meet world demand(130). Recent work introduced the *pfa* genes from the marine bacteria *Moritella Marina* into *E. coli* treated with the antibiotic cerulenin resulted in production of docosahexaenoic acid (DHA), a long-chain PUFA(131). Additional modification by deletion of *fabH* increased DHA production 6.5-fold compared to WT without antibiotic treatment, which was hypothesized to be due uncoupling the canonical fatty acid synthesis pathway and shunting resources to *pfa.* Unfortunately, eliminating the *fab* pathway also resulted in decreased growth, highlighting the need for further design cycles hopefully resulting in acceptable levels of DHA and growth. DHA is known to be toxic to bacteria, resulting in increased lag time, and thus it is unclear where the true origin of production inefficiency is located(132). Eicosapentaenoic acid biosynthesis genes from the marine bacteria *Shewanella putrefaciens* has also been recombinantly produced in *E. coli(133)*, as well as the Δ5-desaturase from *Bacillus subtilis*, resulting in production of double unsaturated fatty acids(134). The rarity of PUFAs biosynthesis in nature has limited bioproduction efforts.

Bioproduction of BCFAs in *E. coli*

Branched chain FAs (BCFAs) are also value-added compounds, primarily in the fuel sector. BCFAs exhibit lower freezing and cloud points, and better flow control(135), indicating potential for their use in higher tolerance applications such as jet engine fuel(136). BCFA biosynthesis pathways are very rare in bacteria, and thus the bioproduction of BCFAs is still in its infancy(137). Machida et al. conduct a phylogenetic analysis of related protein structures and

were able to identify 15 genomes which contained similar genes, several from *Rhodococcus* species, which warrant future observation(137). Another study using *Micrococcus luteus*, a species from phylum *Actinobacteria(138)* identified the BCKD complex of genes responsible for BCFA synthesis, and experimentally confirmed culture conditions leading to increased production of the FA. It remains to be seen whether production can be increased by recombinant expression in *E. coli.* Even more critically, the effects of novel phospholipid compositions on the metabolism of the cell are not well understood and will need to be characterized.

1.2.5 Conclusion and Future Directions

It is exciting to work outside of model organisms in biology. Bioprospecting for genetic information carrying enhanced traits can, and often does, result in real treasure. However, when the time comes for commercial application, or even basic biomolecular characterization of a novel protein product, we often use *E. coli*. It should come as no surprise that during the most recent push for the creation of microbial factories we continually turn to *E. coli(109).* Using a well understood model organism makes the most sense for recombinant protein production, especially with an end goal of human consumption or administration(139). Even after decades of continually refining the process in *E. coli*, successful, high yield of a target protein is often still entails a "trial and error" methodology(140). This is precisely where the utility of large-scale expression profiling can yield results. Properly designed differential expression analysis, in tandem with other -omics techniques, has been shown to be crucial for identifying key metabolic processes linking recombinant protein expression and environmental conditions(141).

Using *E. coli* as a chassis for bioproduction of fatty acids at first makes less sense. *E. coli* only produces a subset of the fatty acid compositions. Given the recent advances for genome editing *Mycobacterium spp.* which produces a more diverse set of FAs, why not use

Mycobacterium smegmatis? The answer lies hidden in the monumental amount of work that goes into producing high expression of recombinant protein. The design-build-test-learn cycle is often conducted as an iterative trial-and-error process with single changes at each step, covering genes of interest and chassis modifications(142). This can take an extremely long time: It took Dupont 15 years and 575 person-years to develop and refine the production process for 1,3-propanediol for commercial viability(143). When there is such a wide space of possible alterations, and the stakes for a company's continued success are so high, it makes sense to reduce as many unnecessary complications as possible.

Moving forward, fatty acid production in *E. coli* will require a systems-biology approach to reduce time to product. This entails integrating top-down retrospective analysis to help identify putative targets overexpression with standard molecular interrogation of phenotype. Using bulk RNAseq analysis alongside overexpression of target proteins, for instance, generates expression data linking individual genetic changes with their effects on larger units of metabolism, such as KEGG modules or pathways(144). This is especially important when developing the *E. coli* chassis for bioproduction of novel phospholipid compositions, which will likely perturb metabolism to a great degree. This methodology has recently been implemented to identify off-target effects of dCas9 overexpression(145). Using RNAseq and ChIP-seq identified hundreds of differentially expressed off-target genes, and upstream binding to 37 genes without single-guide RNA. Studies such as this one provide a framework for implementation in the development of *E. coli* as an efficient chassis for bioproduction of native and novel fatty acids.

1.3 Development of Rhodococcus opacus as a chassis for lignin valorization and bioproduction of high-value compounds

1.3.1 Abstract

The current extraction and use of fossil fuels has been linked to extensive negative health and environmental outcomes. Lignocellulosic biomass-derived biofuels and bio-products are being actively considered as renewable alternatives to the fuels, chemicals, and materials produced from fossil fuels. A major challenge limiting large-scale, economic deployment of second-generation biorefineries is the insufficient product yield, diversity, and value that current conversion technologies can extract from lignocellulose, in particular from the under-utilized lignin fraction. *Rhodococcus opacus* PD630 is an oleaginous gram-positive bacterium with innate catabolic pathways and tolerance mechanisms for the inhibitory aromatic compounds found in depolymerized lignin, as well as native or engineered pathways for hexose and pentose sugars found in the carbohydrate fractions of biomass. As a result, *R. opacus* holds potential as a biological chassis for the conversion of lignocellulosic biomass into biodiesel precursors and other value-added products. This review begins by examining the important role that lignin utilization will play in the future of biorefineries and by providing a concise survey of the current lignin conversion technologies. The genetic machinery and capabilities of *R. opacus* that allow the bacterium to tolerate and metabolize aromatic compounds and depolymerized lignin are also discussed, along with a synopsis of the genetic toolbox and synthetic biology methods now available for engineering this organism. Finally, we summarize the different feedstocks that *R. opacus* has been demonstrated to consume, and the high-value products that it has been shown to

produce. Engineered *R. opacus* will enable lignin valorization over the coming years, leading to cost-effective conversion of lignocellulose into fuels, chemicals, and materials.

1.3.2 Introduction

In comparison to estimated pre-industrial levels (circa 1700 CE), the current global atmospheric CO_2 concentration has increased over 100 parts per million (ppm) and is now stably maintained over 400 ppm, with three-quarters of that change occurring after 1960 due to fossil fuel emissions (146). This increase in the CO_2 concentration has already contributed to a small, but significant, rise in global average temperatures, and will lead to even greater increases in the future. Climate change can lead to decreasing crop yields and seed quality (146), facilitate sea level rise, promote destructive extreme weather events, and cause spikes in energy usage as increasingly frequent severe weather can cause unscheduled shutdown/startup cycles (147). Addressing greenhouse gas-driven climate change will require a complex, multi-tiered approach towards a more carbon-neutral world, including a greater usage of biofuels in the transportation industry and more sustainable chemical and material synthesis.

Biofuels or bio-products, derived from biological sources (i.e., biomass) rather than petroleum, are not a new concept. The inventor of the diesel engine advocated for farmers generating their own vegetable-oil fuel in areas lacking a consistent source of petroleum (148). First-generation biofuels and bio-products are derived from food crops like corn, soy, palm, and sugarcane (148-151). While these biofuels and bio-products have the potential to mitigate CO_2 emissions associated with fossil fuels (148, 150, 151), they are economically and environmentally problematic: energy uses related to irrigation, fertilizer production, cultivation, and transportation are significant, and global demand for food outweighs the supply of arable land (149, 151). Other

sources of biomass, like lignocellulose, are readily available as by-products of the agriculture and forestry industries. Moreover, the dedicated cultivation of properly selected and/or engineered species as sources of lignocellulose can be achieved in a wider distribution of climate and soil conditions with reduced water and fertilizer requirements compared to first-generation sources (152). However, lignocellulose does require more complex processing to produce a second-generation biofuel or bio-product. The structural component of lignocellulose, lignin, provides a particular challenge as it is a complex aromatic macromolecule that evolved to resist degradation (153). Complete utilization and upgrading of lignin is critical for economic viability of second-generation biorefineries.

The processing of lignocellulosic biomass can be split up into two steps: (1) depolymerization in which the polymers within lignocellulose (e.g., cellulose, hemicellulose, and lignin) undergo cleavage reaction producing their respective subunits; and (2) upgrading in which these subunits are converted into a value-added product. Both steps of biomass processing can be achieved through diverse biological, thermochemical, and catalytic processes. However, the generation of a single value-added product requires upgrading followed by extensive chemical separations, or the use of biological catalysts (i.e., a microbe) that can funnel and convert many different lignocellulose-derived substrates into a single product. *Rhodococcus opacus* PD630 (hereafter *R. opacus*) has been identified as a potential biological chassis for the funneling and conversion of lignocellulose-derived substrates to lipids, a biofuel precursor. *R. opacus* has a natural tolerance to toxic aromatic compounds found in the lignin fraction of lignocellulose, an ability to increase this tolerance through adaption, and numerous catabolic pathways for consumption of both carbohydrates and aromatics, making it an ideal candidate to address the challenges of biomass conversion (154). *R. opacus* has also shown the ability to accumulate up to

~80% of its cell dry weight in lipids, such as triacylglycerols (TAGs), under certain growth conditions. These lipids can then be converted to biodiesel via a transesterification reaction (155, 156). In this review, we focus on discussing the important role of lignin valorization in regard to the viability of second-generation biorefineries, summarize different lignocellulose depolymerization methods, and examine *R. opacus*' potential for the conversion of biomass breakdown products into diverse fuels and chemicals.

1.3.4 Why bio-products?

Industrial oil drilling began in the mid-19th century, and in 2015 there was an assessed 35.2 billion barrels of proven oil reserves in the United States, with an estimated 3.4 billion barrels produced domestically that year (157). While predictions of "peak oil" made over the past two decades have been overly pessimistic, it is not unreasonable to predict that recovery of global oil reserves will, at some point, become economically unfeasible (158). Thus, an alternative renewable source of energy and substitutes for products derived from petroleum will be required in the future. Lignocellulosic biomass represents one of the few sources of renewable carbon, while renewable energy (e.g., electricity) can be generated from other renewable energy sources such as wind and solar energy. Biorefineries can convert biomass into a range of products by employing integrated catalytic, thermochemical, and biological conversion processes that efficiently utilize the carbon and energy stored in that biomass (159, 160). A future that does not rely on fossil resources will involve renewable electrical production paired with the generation of biomass-derived products.

The last two decades of research in carbohydrate conversion techniques have witnessed successful biofuel and biochemical production, but the conversion of the lignin fraction of biomass

has been less explored (161). Federal regulation and clean energy initiatives are targeting production rates of 79 billion liters per year of second-generation biofuels by 2022, and the quantity of lignin remaining after the sugar fraction is fermented to reach that target could be as high as 62 million dry tons annually (162-164). Increasing lignin utilization would not only help offset the environmental impact of biomass refinement, but also drastically increase the economic feasibility of the biorefinery. Second generation biofuels and value-added bio-products derived from biomass represent versatile end products, but the future commercial viability of biorefinery products depends on efficient use of both the carbohydrate and lignin substrates (165).

1.3.5 What is lignin?

Lignin is a complex and heterogeneous macromolecule composed of cross-linked aromatic monomers and imparts a rigid or "woody" characteristic of plants that helps provide structural support and limit degradation of polysaccharides. The molecular structure of lignin polymers is primarily derived from *p*-coumaryl, coniferyl, and sinapyl alcohols and corresponds to *p*-hydroxyphenyl, guaiacyl, and syringyl monolignol units, respectively; however, there are a variety of other units which occur less frequently (153, 166). The composition and relative abundance of each type of monolignol varies species-to-species, genotype-to-genotype, across tissue types, between cell wall layers, across different development stages, and as a function of environmental factors (167). The monolignols form several inter-monomer linkages, most commonly aryl ether bonds (e.g., β-O-4, α-O-4, and 4-O-5) (168). The extremely diverse and variable molecular structure of lignin makes commercial degradation difficult and widespread utilization of lignin challenging; however, effective utilization of lignin is necessary for lignocellulose conversion profitability (169).

1.3.6 Thermochemical and catalytic conversion of lignin

Research has been conducted to develop conversion technologies that deconstruct lignin, in particular lignin generated as a by-product of papermaking and biomass-derived carbohydrate fermentation, for the production of renewable fuels, chemicals, and materials (170). Lignin's inherent recalcitrance toward deconstruction makes it difficult to depolymerize for industrial purposes (164, 171). Transforming lignin into higher value products is further complicated by its structural diversity and the high propensity of its intermediates to engage in secondary reactions. An identical lignin conversion process can generate different distributions of compounds depending on the chemical and molecular structure of the lignin feedstock. Additionally, each type of conversion technology has numerous processing conditions that determine the product phase (i.e., solid, liquid, or gas), composition, and application, as well as other conversion performance metrics (e.g. product yield, productivity, selectivity, and composition). For example, a liquid fuel product can easily be derived from lignin using pyrolysis, which exposes the lignin to high temperatures in the absence of oxygen (172). However, the chemical composition of this lignin-derived pyrolysis oil has such a wide distribution of compounds that it has little to no utility for chemical production (173).

Thermochemical and catalytic conversion technologies for lignin valorization primarily include pyrolysis, hydrothermal liquefaction (HTL), gasification, oxidative cracking, hydrogenolysis, and solvolysis (Table 1) (170, 174, 175). These lignin conversion technologies generate gas, liquid, and/or solid breakdown products through numerous complex reactions. Due to differing process conditions defining these conversion technologies, certain reaction pathways are favored, which alters the yield and composition of the breakdown products. Catalytic technologies (i.e., oxidative cracking, hydrogenolysis, and solvolysis), which have been reviewed

in detail by Zakzeski et al, provide a promising avenue to convert lignin selectively into its constituent monomers or monomer derivatives (175). Aromatic carbon-oxygen bonds in aryl ether inter-monomer linkages, which comprise 50-60% of the inter-monomer linkages of lignin (176), represent a potential macromolecular "weak" point that could prove an effective target for selective depolymerization. In this case, a catalyst not only facilitates cleavage of specific bonds along the lignin chains but also allows the cleavage to occur with a lower energy input, reducing the occurrence of secondary reactions with higher activation energies. The resulting product mixture would therefore have a much narrower distribution of aromatic compounds, which may be more amenable to cost-effective chemical separation and/or downstream upgrading.

Table 1.**1. Summary of thermochemical and catalytic technologies for lignin conversion (170).**

Technology	Main Product	Product Application	Process Notes
Gasification	Syngas (gas)	Production of energy, hydrogen & methanol (methanol synthesis); alkanes (Fischer-Tropsch); isobutane (isosynthesis); ethanol (fermentation & catalysts); aldehydes & alcohols (oxosynthesis)	Performed under high temperatures (>700°C); can involve the addition of water & catalyst
Fast pyrolysis / Hydrothermal liquefaction (HTL)	Bio-oil (liquid)	Production of energy and various liquid fuels (e.g., bio-gasoline) by catalytic upgrading	Performed at 250 to 700°C; can involve the addition of water (HTL), hydrogen (hydropyrolsis) & catalyst (catalytic pyrolysis)
Torrefaction / Slow Pyrolysis	Bio-char (solid)	Used as a more-optimized solid fuel for combustion	Performed at 200 to 350°C
Solvolysis	Soluble lignin fragments (liquid)	Phenolics & alkyl-phenolics	Two main categories: (A) acid- & base-catalyzed depolymerization (B) supercritical solvent depolymerization
Hydrogenolysis	Soluble lignin fragments (liquid)	Phenolics & alkyl-phenolics	Hydrogen donor (e.g., hydrogen gas, alcohol, or acid) & a catalyst can be used to cleave linkages
Oxidative cracking	Soluble lignin fragments (liquid)	Aromatic aldehyde, ketones & carboxylic acids	Linkages in lignin can be cleaved by an oxidant (e.g.,

			air & hydrogen peroxide) & a catalyst

1.3.7 Biological and hybrid conversion of lignin

Although catalysts can provide a route toward selective lignin depolymerization chemistry, thermo-catalytic processing of lignin often results in a product mixture that still requires extensive chemical separations. Additionally, lignin depolymerization products are generally limited to aromatic- and phenolic-derivatives. There has been significant research studying the application of enzymes and various microorganisms as a more selective and facile method of lignin depolymerization, as discussed in a recent report (177). In general, biological systems require mild conditions that avoid costs associated with the use of high temperatures and high pressures. However, only a few bacteria (e.g., *Streptomyces* spp., *Rhodococcus* spp., and *Nocardia* spp.) and brown/white-rot fungi (178, 179) have an ability to depolymerize lignin, and their lignin depolymerization rate is too low to be useful on an industrial scale (171, 180).

To overcome these challenges, researchers have adopted a hybrid conversion approach which combines the best attributes of thermo-catalytic and biological conversion technologies (181). In a hybrid conversion approach, a thermo-catalytic conversion process with advantageous reaction kinetics and conversion is applied for the initial lignin depolymerization. Downstream, microbial conversion and funneling of the depolymerized lignin breakdown products (LBPs) to a value-added product then occurs with advantageous selectivity (182). There are numerous aromatic catabolic pathways in various microbes which can be harnessed into a "biological funnel" by converting the heterogeneous substrates generated during depolymerization into common metabolic intermediates (e.g., protocatechuate and catechol) (183). These intermediates undergo

further conversion to central metabolites (e.g., acetyl-CoA) that can be utilized to produce target compounds at a high selectivity.

Hybrid conversion technologies have been implemented, but they have almost exclusively focused on sugar utilization (184, 185). For example, cellulosic technologies can consist of thermochemical polysaccharide depolymerization (e.g., acid hydrolysis (186) or production of pyrolytic sugars (187)) and biological conversion of the resulting monosaccharides into ethanol or other products (184). Recent work has begun to shift the focus from sugar fermentation to lignin utilization, with most research concentrating on using lignin model compounds to characterize aromatic degradation pathways (188-190) and bioconversion abilities (191, 192). Demonstrations of an integrated thermochemical process with an aromatic-metabolizing microbial catalyst using pretreatment liquors have been performed, but these feedstocks frequently contain only a portion of the original lignin content, as the pretreatment process has been optimized for sugar release via enzymatic hydrolysis rather than maximizing lignin conversion (183, 193, 194). For an effective lignin hybrid conversion process, the upstream thermochemical or catalytic depolymerization process must meet the following requirements: (1) production of aqueous soluble LBPs; (2) optimization of lignin conversion for yield and selectivity toward the preferred substrates for microbial growth and utilization; (3) minimal generation of inhibitor compounds; and (4) a process configuration and condition that is compatible with an economical, sustainable, and large-scale design. To this end, multiple lignocellulosic biomass pretreatment techniques have been tested, demonstrating the potential of hybrid conversion processes (193-197).

1.3.8 Why *Rhodococcus opacus* PD630 for hybrid conversion of lignin?

Rhodococcus opacus PD630 has been identified as a candidate biological catalyst for the conversion of both the carbohydrate and lignin fractions of lignocellulose into valuable products. *R. opacus* was originally isolated from soil collected near a gas works plant by enrichment on phenyldecane as a sole carbon source (156). *R. opacus* possesses extensive catabolic pathways for both sugars and aromatics and can tolerate inhibitory compounds found in depolymerized biomass (e.g., phenolics and furfural) (198). The ability to metabolize aromatic compounds is shared by a number of microorganisms and is likely a common evolutionary trait due to the prevalence of lignin in natural environments. Many of the aromatic compounds that *R. opacus* is known to metabolize can be found in LBPs (154, 199-203). In addition to lignin model compounds (e.g., 4-hydroxybenzoate, benzoate, phenol, vanillate, guaiacol, and trans-*p*-coumaric acid), *R. opacus* has been shown to degrade depolymerized kraft lignin (204), alkali-treated corn stover (205, 206), alkali-treated poplar wood (179), and switchgrass pyrolysis oil (197). Through adaptive evolution, *R. opacus* has been further evolved to more efficiently degrade phenol, syringaldehyde, and aromatic mixtures (154, 198, 199). *R. opacus* has also been engineered using exogenous genes expressed on plasmids to degrade cellulose, arabinose, and xylose (207-209). It is therefore, through native or engineered means, able to tolerate and utilize a variety of typically toxic lignin-derived compounds, in addition to sugars.

Unlike most bacteria that store carbon as polyhydroxyalkanoic acids (PHAs), *R. opacus* stores carbon as energy-rich triacylglycerols (TAGs) (203, 210). Acetyl-CoA is the product of diverse catabolic pathways in *R. opacus*, including glycolysis, the Entner–Doudoroff pathway, and aromatic degradation pathways (e.g., β-ketoadipate pathway), and it is a key precursor in TAG biosynthesis. Under nitrogen limitation, the non-limiting essential nutrient (i.e., carbon) is stored

as TAGs in *R. opacus*, accumulating up to ~80% of its cellular dry weight when cultured on gluconate (156, 211). On aromatic compounds, lipid production is reduced, but it can still reach up to 44% of cellular dry weight in TAGs under nitrogen-limiting conditions (154). *R. opacus* can also synthesize branched-chain and odd-numbered fatty acids that are necessary for next generation biofuels. Shifting lipid storage in *R. opacus* to these compounds would make it an even more valuable production strain (212).

Other organisms have been proposed for lignin conversion, but *R. opacus* has demonstrated higher or equal rates of aromatic degradation and tolerance compared to other species. For example, a phenol-adapted *Pseudomonas putida* strain had a maximum phenol degradation rate of ~12 mg phenol/L/hr when grown at its maximum tolerated phenol concentration of 1 g/L (213). *Bacillus brevis* previously claimed the highest phenol tolerance and utilization when cultures grew at concentrations up to 1.75 g/L phenol, and demonstrated a maximum degradation rate of ~20 mg/L/hr (214). Adapted *R. opacus* strains were able to grow at 2 g/L phenol, and demonstrated a maximum degradation rates of ~21-22 mg/L/hr (199).

In summary, *R. opacus* is an ideal candidate for the hybrid approach of lignocellulose utilization because of its high tolerance to aromatic compounds, its capacity to utilize a wide variety of substrates (both carbohydrates and lignin), and its ability to accumulate lipids. These traits are uniquely-suited to handling the aromatic mixtures produced by lignin depolymerization processes and metabolizing them into valuable compounds. Additionally, *R. opacus* is amenable to adaptive evolution to improve the tolerance and growth rate on aromatic and lignin substrates (154, 198, 199). These natural characteristics, along with a growing toolbox of genetic tools, make *R. opacus* an ideal organism for lignin valorization (215).

1.3.9 Genetic and metabolic characteristics of *R. opacus*

Aromatic degradation in *R. opacus* is facilitated by a high-flux β-ketoadipate pathway that produces acetyl-CoA (**Error! Reference source not found.**) (216).

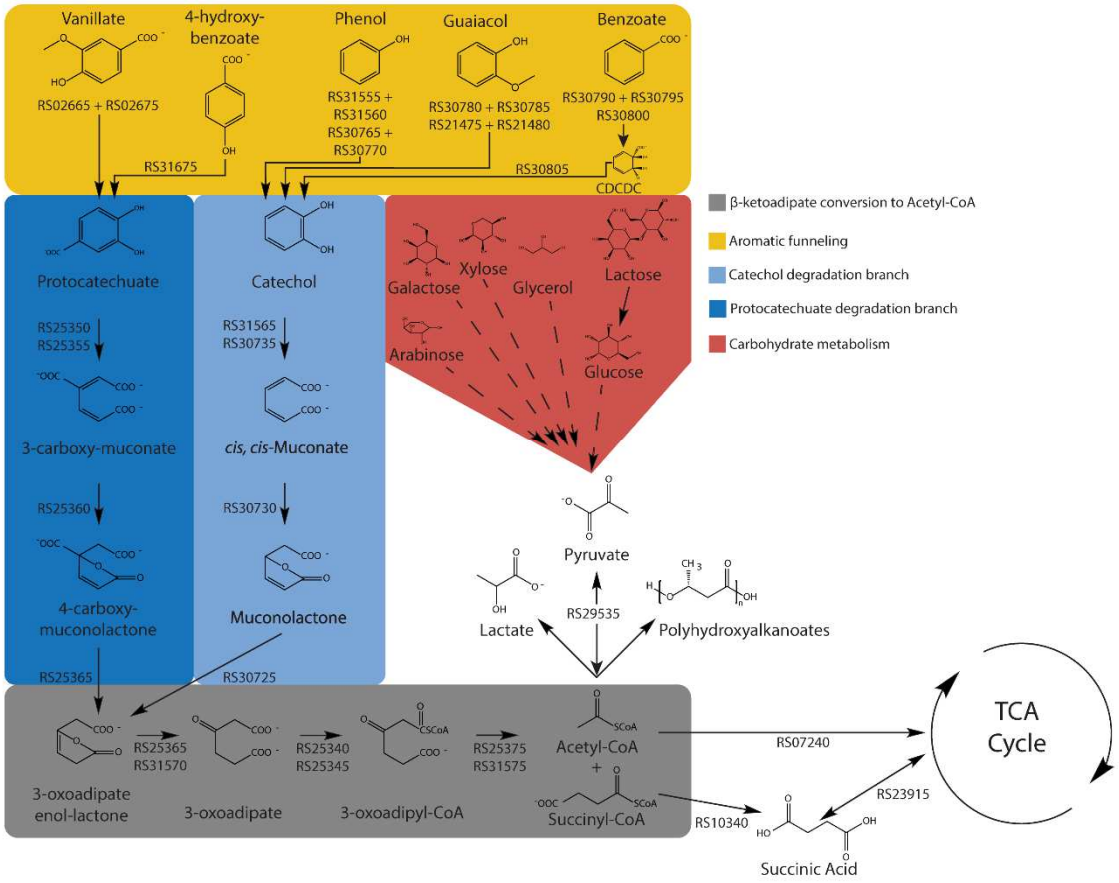


Figure 3. 2. **Aromatic degradation and carbon metabolism in R. opacus.** *R. opacus* genes involved in reactions are listed. Dashed arrows represent multiple intermediate steps not shown. Xylose and arabinose consumptions occur via engineered pathways.

As acetyl-CoA is the precursor molecule for many biochemicals, *R. opacus* is thus well suited for chemical production based on an aromatic feedstock. Additionally, glucose metabolism in *R. opacus* exclusively utilizes the Entner-Doudoroff pathway, enabling simultaneous utilization of phenol and glucose. This lack of catabolite repression means that *R. opacus* can effectively use

both the carbohydrate and lignin fractions of lignocellulosic biomass with reduced fermentation times and increased productivities. While sugar and phenol metabolisms are independent, *R. opacus* degrades aromatic compounds in a preferential order (154, 217). It is unclear what is driving this preferential consumption, but it may result from variations in enzyme and transporter activities, or transcriptional-level regulation.

Aromatic compounds entering the cell generally first undergo preliminary degradation to either protocatechuate or catechol before being metabolized through the β-ketoadipate pathway. Mechanistically, this import into the cell and pre-processing occurs via specialized aromatic transporters and funneling enzymes, which have been identified using transcriptomics and proteomics (154, 179, 199). Henson et al. identified three aromatic-associated transporters: one specific to phenol, one associated with both phenol and vanillate, and a promiscuous transporter associated with phenol, vanillate, benzoate, and guaiacol (154). Several funneling enzymes have also been identified, including those which convert vanillate and 4-hydroxybenzoate to protocatechuate, as well as those which convert phenol, guaiacol, and benzoate to catechol (154).

Advantageous mutations and transcriptional changes have been identified in *R. opacus* that could be future targets for additional growth optimization through forward engineering. Genes for enzymes involved in oxidation-reduction reactions underwent changes in multiple strains adapted for improved growth on one or more phenolic compounds, including cytochrome ubiquinol oxidase subunit I and superoxide dismutase (154, 199). The fact that functionally equivalent mutations occurred in multiple aromatic-adapted strains suggests their link to improved aromatic tolerance and utilization. For example, decreasing the activity of superoxide dismutase, as demonstrated in these mutated strains, may allow the cells to increase oxidizing equivalents, which

are necessary to degrade highly-reduced aromatic rings. Transcriptomic analysis of adapted strains also identified increased expression of aromatic transporters, which correlated with increased phenolic tolerance and utilization. These and other changes identified in the genome and transcriptome of adaptively-evolved strains could be replicated in a rationally-engineered strain to fine-tune its growth on and tolerance to aromatic-rich substrates.

Additional potential targets for strain optimization in *R. opacus* are its nine endogenous plasmids (2 circular and 7 linear plasmids; a combined total of 0.79 Mbp). These plasmids have been posited to act as a hyper-recombinational gene storage strategy in which infrequently-used catabolic genes are stored on plasmids as a failsafe against rarely-encountered compounds (e.g., nitrophenolates and polycyclic and/or halogenated aromatics) which may be present in the environment (218-223). If genes located on a plasmid are found to be regularly useful, they can undergo recombination with the 8.38 Mbp circular chromosome and become permanent components of the genome. This strategy has been observed in strains of related *Actinomycetales*, where stored genes provide the catabolic versatility to degrade a larger array of organic compounds (224, 225). Additionally, genes located on plasmids, particularly if they are duplicates not subject to evolutionary conservation, can collect mutations more rapidly than those genes in the chromosome, allowing for improved adaptive capacity. For potential industrial use with a relatively well-defined and consistent feedstock, tailored strains of *R. opacus* may benefit from the selective removal of some of these plasmids, as previously adapted strains that exhibited improved growth profiles on phenolic compounds underwent large deletions or complete loss of plasmids (154). Plasmid removal under selective pressure may be driven by a reduced metabolic burden. Strategically, intentional plasmid curing would best be employed when cells are cultured on a

defined range of carbon sources, where trading catabolic potential for improved growth rate is an acceptable risk.

1.3.10 Tool and technique development for R. opacus engineering

As *R. opacus* is a non-model bacterium, the available genetic tools and techniques for engineering this organism were relatively sparse until recently. The genetic toolbox available for *R. opacus* has expanded to include a reference genome (203, 210), plasmid backbones for gene overexpression, and promoters for tunable gene expression (Table 2). Furthermore, methods for performing gene knock-outs and knock-ins, modulating and quantifying gene expression, and extracting intracellular products via viral lysis have all been demonstrated. We summarize the most prominent genetic tools and techniques developed for several *R. opacus* strains.

A common element used for gene overexpression is the replicating plasmid backbone. There are a number of heterologous plasmids isolated from other *Actinomyces* spp. that have been demonstrated to replicate stably in *R. opacus* (Table 2). While they have several names, these plasmid backbones can be grouped into five primary categories. The pAL5000-based plasmid group, consisting of short and long variants, is derived from *Mycobacteria* spp. and has been demonstrated to have 3 to 11 copies per chromosome, dependent on the variant (218, 226, 227). The pNG2, pGA1, and pSR1 plasmid groups are all ancestrally related and derived from cryptic *Corynebacterium* spp. plasmids that replicate through rolling-circle amplification (228-232). A BioBrick-compatible version of pSR1 (pSRKBB) has recently been developed for easy cloning (233). The pAL5000 and pNG2 backbones have been demonstrated to be compatible, allowing co-maintenance of two heterologous plasmids (218). Finally, pB264, which is derived from an endogenous *Rhodococcus sp.* B264 plasmid, has ~8 copies per chromosome and is easily curable

from the cell once antibiotic selection pressure is removed (218, 234). To ensure that heterologous plasmids are stably maintained within a cell, selection is required, most frequently in the form of an antibiotic resistance marker. Several of these markers have been demonstrated and optimized for selection in *R. opacus* (Table 2), though there may be room for refinement with regards to stable maintenance concentrations (233). A recent study employed single-cell fluorescence to demonstrate that the concentration of kanamycin sufficient for selection was not the ideal concentration for plasmid function maintenance. The commonly-used concentration of 50 μg/mL led to a bimodal population of cell fluorescence, with fewer than half of the cells demonstrating fluorescent reporter expression when analyzed via flow cytometry (233). Increasing the concentration of kanamycin to 250 μg/mL led to a majority of cells expressing the fluorescent reporter.

Table 1.2. List of genetic parts demonstrated in *R. opacus*, including plasmid backbones, selection markers, promoters, and recombination-related parts.

Part Type	Name	Properties/Notes	Source
Plasmid backbones	pAL5000 (short)	Other names: pXYLA and pNV18; ~11 copies per chromosome	(207, 218, 235-239)
	pAL5000 (long)	Other names: pJAM2 and pJEM; ~ 3 copies per chromosome	(208, 218, 235, 238)
	pNG2	Derived from *Corynebeacterium* spp.; ~10 copies per chromosome	(227, 229)
	pGA1	Derived from *Corynebeacterium* spp.	(208, 232, 240, 241)
	pSR1	Derived from *Corynebeacterium* spp.	(230, 231, 233)
	pB264	Derived from *Rhodococcus* sp. B264; curable; ~8 copies per chromosome	(218)

Selection markers	Kanamycin	50 μg/mL (selection) 250 μg/mL (plasmid function maintenance)	(230, 233, 242)
	Gentamicin	10 μg/mL	(227)
	Spectinomycin	100 μg/mL	(227)
	Thiostrepton	1 μg/mL	(242)
	Chloramphenicol	34 μg/mL	(243, 244)
	Hygromycin B	50 μg/mL	(218)
	SacB	Negative selection; sensitizes cell to sucrose	(245, 246)
Promoters	pTipA	Inducible with thiostrepton	(242, 247, 248)
	pAcet	5x inducible with acetamide	(243, 249)
	pBAD	59x inducible with arabinose	(249)
	pTet	67x inducible with anhydrotetracycline (aTc)	(249)
	pLPD06740	247x inducible with phenol	(249)
	pLPD06575	Inducible with phenol	(249)
	pLPD06699	39x inducible with phenol, protocatechuic acid, sodium benzoate, 4-hydroxybenzoate, vanillate, and guaiacol	(249)
	pLPD06568	80x inducible with phenol, sodium benzoate, and guaiacol	(249)
	pLPD03031	18x repressible with ammonium	(249)
	IGRI' & IGRIV'	Inducible with 2,4-dinitrophenol (DNP)	(245)
Recombinases	Che9c60	GC-rich homologue of RecE	(218, 224, 250)

	Che9c61	GC-rich homologue of RecT	(218, 224, 250)
Neutral sites	ROCI-2	*R. opacus* chromosomal locus	(218)
	ROCI-3	*R. opacus* chromosomal locus	(218)
	ROP8I-1	*R. opacus* endogenous plasmid 8 locus	(218)

For promoters inducible with multiple chemicals, the highest published fold-change for a single compound is reported.

The ability to readily transform a bacterium with heterologous DNA is critical if it is to be a platform organism. *R. opacus* is transformable through multiple methods, including conjugation and electroporation. Conjugation requires a plasmid containing an origin of transfer (OriT) and a bacterial strain capable of conjugating with the strain of interest to horizontally transfer the plasmid (251). Both *E. coli* DH5α-pKOS111-47 and *E. coli* S17.1 have been used as conjugative helper strains with *R. opacus* (243, 245). Electroporation, wherein a pulse of electricity creates pores in the cellular membrane, can facilitate uptake of plasmid DNA in *R. opacus* at a reported efficiency of ~10^5 CFUs/μg DNA (242, 252).

Successful cellular engineering requires the use of well-characterized genetic parts for predictable gene expression, and in non-model organisms, parts are often borrowed from related organisms (253). One core component needed for reliable gene expression is the promoter, which drives gene transcription. A number of studies have utilized constitutive promoters from related gram-positive *Actinomycetales* (e.g. *Mycobacterium* spp. and *Streptomyces* spp.) or from genetically-distant bacteria, such as gram-negative *E. coli*, for heterologous gene expression (207, 208, 227, 233, 241). When performing metabolic engineering, however, a number of different promoters of varying strengths are required to balance the expression of multiple genes in an enzymatic pathway for optimal product titers (253, 254). An alternative to using borrowed

promoters is the creation of a *de novo* constitutive promoter library, where many promoters of varying strengths are developed and characterized. Using a fluorescent reporter, a constitutive promoter library spanning a 45-fold change in fluorescent output from weakest to strongest promoter was generated for *R. opacus* (218). Performing initial optimization of an enzymatic pathway combinatorically with a range of constitutive promoters, however, can be time-consuming and costly in non-model organisms.

An alternative to constitutive promoters is tunable promoters whose expression is induced or repressed relative to the concentration of a specific compound. Table 2 summarizes the inducible and repressible promoters that have been demonstrated in *R. opacus*. Of particular interest to the goal of using *R. opacus* for the conversion of lignin to lipids are the aromatic- and ammonium-responsive promoters. The aromatic promoters (pLPD06740, pLPD06575, pLPD06699, pLPD06568) are differentially induced in the presence of a variety of aromatic compounds, including some found in depolymerized lignin, and could be employed in metabolic engineering related to aromatic catabolism (249). pLPD03031 is a promoter that is repressed in the presence of ammonium, which can be used as a sole nitrogen source in *R. opacus*, and turns on when ammonium is depleted (249). As nitrogen starvation triggers lipid accumulation in *R. opacus*, this promoter could be used to modulate lipid pathways under lipid accumulating conditions (155, 199). Combining the aromatic- and ammonium-responsive promoters into genetic circuits could lead to dynamic regulation, which has been shown to increase final product titers through reductions in metabolic burden (174). Furthermore, pLPD03031 has been employed to create a cellular timer designed to activate at specific points in the cellular growth cycle, dependent on the initial ammonium concentration in the culture (249).

In addition to expressing a gene construct on a plasmid, heterologous expression can be achieved by integrating the DNA into the genome of the organism, where it can be stably maintained. To date, genome modification in *R. opacus* has been performed through both single- and double-crossover homologous recombination (218, 243, 245, 250). Two methods utilizing single-crossover recombination, combined with the conjugative transfer of a donor plasmid via an *E. coli* helper strain, have been described for *R. opacus* (243, 245). One difficulty with genomic recombination in *R. opacus*, however, is that it often results in illegitimate integration, wherein the integration cassette is inserted at an incorrect location or the entire plasmid is integrated. This is a common issue in other actinobacteria, such as *Mycobacterium tuberculosis* and *Rhodococcus fascians*, and can be overcome through the heterologous expression of helper recombinases (224, 255). In *R. opacus*, a pair of bacteriophage recombinases, Che9c60 and Che9c61, have been demonstrated to facilitate double-crossover homologous recombination when donor template is provided via electroporation (218, 224, 250). Ideally, the integration of foreign DNA into the genome would have no adverse effects on cell health, but in practice care must be taken in choosing an integration site. Three neutral sites, or locations that have been demonstrated not to cause a decrease in growth rate when a gene cassette is integrated into them, have been identified in the chromosome and a native endogenous plasmid of *R. opacus* (218).

In addition to the tools that facilitate gene overexpression in *R. opacus*, it may also be desirable to eliminate certain genes. Genetic knockouts through homologous recombination have also been performed to disrupt gene expression. Both single- and double-crossover recombination have been used to knockout and confirm the functional roles of transcriptional regulators, catabolic enzymes, and transporters in *R. opacus* (154, 245). Furthermore, the *sacB* negative selection

marker, which sensitizes the cell to sucrose, has been used for genome engineering in *R. opacus* (245, 246).

Gene knockouts can be informative when investigating gene function, but as permanent modifications, they may be lethal to the cell. As an alternative, a gene's expression can be selectively and temporarily reduced through CRISPR interference (CRISPRi). CRISPRi utilizes a complex comprising a deactivated Cas9 (dCas9) and an engineered small guide RNA (sgRNA) to bind to DNA in a sequence-dependent manner and interfere with the transcriptional machinery, leading to targeted gene repression (256). The most commonly used CRISPRi system is derived from *Streptococcus pyogenes* ($dCas9_{Spy}$), but this system was found to be ineffective in *Mycobacterium tuberculosis*, an *Actinomycetales* species closely-related to *R. opacus* (257). A version of dCas9 sourced from *Streptococcus thermophilus* ($dCas9_{Sth1}$), which was found to be effective in *M. tuberculosis*, has been developed as a repression system for use in *R. opacus* (218). Experimentally, up to 58% repression of a chromosomally-integrated fluorescent protein was observed using this optimized $dCas9_{Sth1}$(218). Tunable gene repression using $dCas9_{Sth1}$ can be used in the future to remodel native metabolic pathways in *R. opacus*.

To quantify changes in gene expression between different growth conditions, stably-expressed reference (or "housekeeping") genes in *R. opacus* have been identified for use with reverse transcription quantitative PCR (RT-qPCR) (258). When ribosomal RNAs (rRNAs) are present in the samples, it was found that the combined use of genes for the ATP-binding subunit, ClpX, of the ATP-dependent Clp protease (PD630_RS25530) and 16S rRNA (PD630_RS01395) provided the best normalization results. If rRNAs are depleted, as is the case in samples prepared for RNA-Seq, the best pair of genes was found to be the same ATP-binding subunit ClpX and the

rRNA small subunit methyltransferase G (PD630_RS37755). Using an appropriate set of reference genes is essential to generate meaningful expression-change data, and these pairs provide this baseline to an array of analyses.

A final technique that has been implemented to improve bioproduction in *R. opacus* is the development of a method for controlled cellular lysis to release intracellular compounds (259). A bottleneck in microbial manufacturing is separating the target product from the cells, and implementing a controlled release strategy could reduce processing costs. A domesticated version of *R. opacus* created through serial culturing was found to be sensitive to a lytic tectivirus (Phage Toil) (259), which can be used to trigger cell lysis at a desired timepoint and can thus serve as a cheap and effective method for releasing products (e.g. TAGs) from *R. opacus* (259).

1.3.11 *R. opacus* as a production host

The production of TAGs and fatty acids in *R. opacus* has been demonstrated on an array of carbon feedstocks (Table 3). When fed kraft lignin (a toxic byproduct of the paper and pulping industry) in combination with laccase (a class of enzymes which oxidize phenolics), *R. opacus* was able to generate 0.145 g/L of lipid (204). A strain of *R. opacus* adaptively-evolved to tolerate higher levels of aromatic compounds, PVHG6, was able to generate 0.13 g/L lipids when provided with five lignin model compounds as sole carbon sources in equal quantities (154). Growth of *R. opacus* on pre-treated corn stover produced 1.3 g/L of lipid (measured as fatty acid methyl esters (FAME)) (205). A xylose-fermenting strain (MITXM-61) was developed by heterologous gene expression, and when it was grown in corn stover hydrolysates (containing 118 g/L initial total sugars), it converted xylose and glucose into 15.9 g/L TAGs (54% of dry cell weight (DCW)) (260). When *R. opacus* was cultured in glucose and glycerol, Suwaleerat et al. observed a

maximum of 2.4 g/L lipids in 10.2 g/L biomass (261). While *R. opacus* grows poorly on glycerol alone (261, 262), this demonstrates that it can be used to enhance lipid production compared to just glucose as a sole carbon source. As glycerol is a byproduct of TAG transesterification, feeding it back to the production strain could reduce overall costs (263). Using adaptive evolution, Sinskey and colleagues generated an MITXM-61 derivative strain, MITGM-173, which was able to grow on up to 160 g/L glycerol (264). Optimized TAG production in this adapted strain occurred with a 1:2:2 mixture of glycerol:glucose:xylose, reaching 13.6 g/L TAGs (51.2% of DCW).

Table 3. Bioproduction by *R. opacus* wild-type and engineered strains on various feedstocks.

Strain	Substrate	Product	Production Value	Reference
R. opacus PD630	glucose:glycerol (7:3)	Carotenoids & lipids	0.99 mg/L & 2.4 g/L, respectively	(261)
R. opacus PD630	Pre-treated Corn Stover	FAME	1.3 g/L	(205)
R. opacus PD630	Glycerol	TAGs	1.4 g/L, 38.4% DCW	(262)
R. opacus PD630 (engineered)	Glucose	Fatty acids	46% DCW	(265)
R. opacus MITXM-61 (engineered)	Corn stover hydrolysates	TAGs	15.9 g/L, 54% DCW	(260)
R. opacus MITGM-173 (evolved)	glycerol: glucose:xylose (1:2:2)	TAGs	13.6 g/L, 51.2% DCW	(264)
R. opacus PD630	Crude Whey	Fatty Acids	45.1% DCW	(266)
R. opacus PD630	Switch Grass (pyrolysis oil)	Lipid	pH 7: 0.078 g/L, 21.9% DCW pH 4: 0.066 g/L, 25.8% DCW	(197)

R. opacus PD630	Kraft lignin (+ laccase)	Lipid	0.145 g/L	(204)
R. opacus PD630	Olive Mill Waste	Lipid	~1.9 g/L, 80% DCW	(267)
R. opacus PD630 PVHG6	phenol:vanillate:4-hydroxybenzoate: guaiacol:benzoate (1:1:1:1:1)	Lipid	0.13 g/L, 44% DCW	(154)
R. opacus PD630 (engineered)	Gluconate, whey	Wax esters	Gluconate: 46% total neutral lipids Whey: NR	(244)
R. opacus PD630	Poplar lignin hydrolysis slurry	Lipid	NR	(179)

FAME: fatty acid methyl ester; DCW: dry cell weight; NR: not reported.

R. opacus can also be used as a platform to produce high value compounds other than lipids. For example, *R. opacus* naturally produces carotenoids, which are prized for their pigmentation and antioxidant properties (261). In a 7:3 ratio of glucose and glycerol, *R. opacus* was able to produce 0.99 mg/L of carotenoids (261). Though *R. opacus* has, at present, a limited pool of demonstrated products (i.e., lipids and carotenoids), there is potential to expand the range, particularly as its genetic toolbox has recently been developed. With rational metabolic engineering, it may be possible to shunt more carbon flux into carotenoid production to improve yields and titers. Similarly, manipulation of enzymes in the β-ketoadipate pathway (**Error! eference source not found.**) could result in the accumulation of high-value intermediates, including *cis, cis*-muconic acid (bioplastic precursor) and succinic acid (food additive). Furthermore, the downstream product of aromatic degradation, acetyl-CoA, can be diverted to produce diverse compounds. Production of these high-value compounds from lignocellulose or

lignocellulose-derived sources has been demonstrated in other bacterial hosts, which provides a guide for engineering *R. opacus* (Table 4).

Table 4. Compounds produced from lignin or lignin-derived sources in selected non-*R. opacus* bacterial hosts.

Strain	Substrate	Product	Production Value	Reference
R. jostii RHA1 (engineered)	Wheat straw	Vanillin	96 mg/L	(268)
P. putida KT2440 (engineered)	Depolymerized corn stover lignin	*Cis, cis*-muconic acid	3.7 g/L	(269)
Corynebacterium glutamicum (engineered)	Depolymerized softwood lignin	*Cis, cis*-muconic acid	1.8 g/L	(270)
Cupriavidus basilensis B-8	Kraft lignin	Polyhydroxyalkanoate (PHA)	319.4 mg/L	(271)
Actinobacillus succinogenes 130Z	Xylose-enriched corn stover hydrolysate	Succinic acid	39.6 g/L	(272)

1.3.112 Conclusion

Biofuels and bioproducts can be produced from lignocellulose to replace or supplement petroleum-based fuels, chemicals, and materials. To improve the economic competitiveness and reduce the environmental footprint of biorefineries, both the carbohydrate and lignin fractions should be utilized. However, due to its recalcitrance, lignin has been an untapped carbon source which is either discarded or burned for process heat. Additionally, aromatic compounds found in depolymerized lignin are toxic to most microbes, presenting a challenge to developing an

economically viable process. To overcome these challenges, we propose a hybrid conversion approach that combines thermochemical/catalytic and biological conversion processes as discussed in this review.

R. opacus is an ideal organism for such a hybrid conversion process due to its ability to tolerate and utilize a wide variety of aromatic compounds found in lignin breakdown products. Additionally, *R. opacus* is oleaginous and can produce high levels of lipids. While *R. opacus* engineering has been limited, recent identification of aromatic degradation pathways and substrate transporters has provided several targets that can be modified for strain optimization. In addition, the toolbox for genetic engineering is under active development, providing methods for gene modification and transcription control. Further engineering will be necessary to increase the tolerance, growth rate, and lipid production of *R. opacus* on depolymerized lignin substrates. Furthermore, *R. opacus* can be engineered to synthesize more valuable chemicals, such as branched-chain fatty acid esters, carotenoids, and *cis, cis*-muconic acid. With advances in both lignin depolymerization processes and rational engineering tool development for *R. opacus*, the coming years will witness rapid progress toward cost-effective conversion of lignocellulose into bio-products.

1.4 Acknowledgements

This work was adapted from a manuscript written in collaboration with Rhiannon R. Carr, Drew M. DeLorenzo, Tayte Campbell, Zeyu Shang, Marcus Foston, Tae Seok Moon, and Gautam Dantas, published in *BMC Biotechnology for Biofuels* (now *BMC Biotechnology for Biofuels and Bioproducts*)(273). The authors thank members of the Moon, Dantas, and Foston research groups for scientific discussions.

Authors' contributions

All authors contributed text to the manuscript. W.E.A. generated the figure. All authors read and approved the final manuscript.

Funding

U.S. Department of Energy [DE-SC0018324 to M.F., G.D., and T.S.M.]

1.4 References

1. Redelings MD, Sorvillo F, Mascola L. Increase in Clostridium difficile-related mortality rates, United States, 1999-2004. Emerg Infect Dis. 2007;13(9):1417-9.

2. Colomb-Cotinat M, Lacoste J, Brun-Buisson C, Jarlier V, Coignard B, Vaux S. Estimating the morbidity and mortality associated with infections due to multidrug-resistant bacteria (MDRB), France, 2012. Antimicrobial Resistance & Infection Control. 2016;5(1):56.

3. Furuya-Kanamori L, Marquess J, Yakob L, Riley TV, Paterson DL, Foster NF, et al. Asymptomatic Clostridium difficile colonization: epidemiology and clinical implications. BMC Infectious Diseases. 2015;15(1):516.

4. Kyne L, Warny M, Qamar A, Kelly CP. Asymptomatic Carriage of Clostridium difficile and Serum Levels of IgG Antibody against Toxin A. New England Journal of Medicine. 2000;342(6):390-7.

5. Belmares J, Johnson S, Parada JP, Olson MM, Clabots CR, Bettin KM, et al. Molecular Epidemiology of Clostridium difficile over the Course of 10 Years in a Tertiary Care Hospital. Clinical Infectious Diseases. 2009;49(8):1141-7.

6. Blaser MJ. Disappearing Microbiota: Helicobacter pylori Protection against Esophageal Adenocarcinoma. Cancer Prevention Research. 2008;1(5):308-11.

7. Gacesa R, Kurilshikov A, Vich Vila A, Sinha T, Klaassen MAY, Bolte LA, et al. Environmental factors shaping the gut microbiome in a Dutch population. Nature. 2022;604(7907):732-9.

8. He Y, Wu W, Zheng H-M, Li P, McDonald D, Sheng H-F, et al. Regional variation limits applications of healthy gut microbiome reference ranges and disease models. Nature Medicine. 2018;24(10):1532-5.

9. Singh RK, Chang H-W, Yan D, Lee KM, Ucmak D, Wong K, et al. Influence of diet on the gut microbiome and implications for human health. Journal of Translational Medicine. 2017;15(1):73.

10. Davenport ER, Mizrahi-Man O, Michelini K, Barreiro LB, Ober C, Gilad Y. Seasonal Variation in Human Gut Microbiome Composition. PLoS One. 2014;9(3):e90731.

11. Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. P T. 2015;40(4):277-83.

12. Chow J, Tang H, Mazmanian SK. Pathobionts of the gastrointestinal microbiota and inflammatory disease. Current opinion in immunology. 2011;23(4):473-80.

13. Penders J, Stobberingh E, Savelkoul P, Wolffs P. The human microbiome as a reservoir of antimicrobial resistance. Frontiers in Microbiology. 2013;4.

14. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome Med. 2016;8(1):51.

15. Buffie CG, Jarchum I, Equinda M, Lipuma L, Gobourne A, Viale A, et al. Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to Clostridium difficile-induced colitis. Infection and immunity. 2012;80(1):62-73.

16. Ling Z, Liu X, Jia X, Cheng Y, Luo Y, Yuan L, et al. Impacts of infection with different toxigenic Clostridium difficile strains on faecal microbiota in children. Scientific Reports. 2014;4(1):7485.

17. Kim S, Covington A, Pamer EG. The intestinal microbiota: antibiotics, colonization resistance, and enteric pathogens. Immunological reviews. 2017;279(1):90-105.

18. McInerny GJ, Etienne RS. Stitch the niche – a practical philosophy and visual schematic for the niche concept. Journal of Biogeography. 2012;39(12):2103-11.

19. Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, et al. Competitive and cooperative metabolic interactions in bacterial communities. Nature Communications. 2011;2(1):589.

20. Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: Networks, competition, and stability. Science. 2015;350(6261):663-6.

21. Bauer MA, Kainz K, Carmona-Gutierrez D, Madeo F. Microbial wars: Competition in ecological niches and within the microbiome. Microb Cell. 2018;5(5):215-9.

22. Welsh RM, Zaneveld JR, Rosales SM, Payet JP, Burkepile DE, Thurber RV. Bacterial predation in a marine host-associated microbiome. ISME J. 2016;10(6):1540-4.

23. Dethlefsen L, Eckburg PB, Bik EM, Relman DA. Assembly of the human intestinal microbiota. Trends in Ecology & Evolution. 2006;21(9):517-23.

24. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011;473(7346):174-80.

25. Cheng M, Ning K. Stereotypes About Enterotype: the Old and New Ideas. Genomics Proteomics Bioinformatics. 2019;17(1):4-12.

26. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. PLOS ONE. 2015;10(5):e0128036.

27. Liang C, Tseng H-C, Chen H-M, Wang W-C, Chiu C-M, Chang J-Y, et al. Diversity and enterotype in gut bacterial community of adults in Taiwan. BMC Genomics. 2017;18(1):932.

28. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011;334(6052):105-8.

29. Checinska Sielaff A, Urbaniak C, Mohan GBM, Stepanov VG, Tran Q, Wood JM, et al. Characterization of the total and viable bacterial and fungal communities associated with the International Space Station surfaces. Microbiome. 2019;7(1):50.

30. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486(7402):222-7.

31. Wilmanski T, Diener C, Rappaport N, Patwardhan S, Wiedrick J, Lapidus J, et al. Gut microbiome pattern reflects healthy ageing and predicts survival in humans. Nat Metab. 2021;3(2):274-86.

32. Koliada A, Moseiko V, Romanenko M, Piven L, Lushchak O, Kryzhanovska N, et al. Seasonal variation in gut microbiota composition: cross-sectional evidence from Ukrainian population. BMC microbiology. 2020;20(1):100.

33. Shaw LP, Bassam H, Barnes CP, Walker AS, Klein N, Balloux F. Modelling microbiome recovery after antibiotics using a stability landscape framework. ISME J. 2019;13(7):1845-56.

34. Wipperman MF, Fitzgerald DW, Juste MAJ, Taur Y, Namasivayam S, Sher A, et al. Antibiotic treatment for Tuberculosis induces a profound dysbiosis of the microbiome that persists long after therapy is completed. Scientific Reports. 2017;7(1):10767.

35. Jernberg C, Löfmark S, Edlund C, Jansson JK. Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. ISME J. 2007;1(1):56-66.

36. Adu-Oppong B, Gasparrini AJ, Dantas G. Genomic and functional techniques to mine the microbiome for novel antimicrobials and antimicrobial resistance genes. Ann N Y Acad Sci. 2017;1388(1):42-58.

37. Lanza VF, Tedim AP, Martínez JL, Baquero F, Coque TM. The Plasmidome of Firmicutes: Impact on the Emergence and the Spread of Resistance to Antimicrobials. Microbiology Spectrum. 2015;3(2).

38. Dolk FCK, Pouwels KB, Smith DRM, Robotham JV, Smieszek T. Antibiotics in primary care in England: which antibiotics are prescribed and for which conditions? Journal of Antimicrobial Chemotherapy. 2018;73(suppl_2):ii2-ii10.

39. D'Costa VM, King CE, Kalan L, Morar M, Sung WW, Schwarz C, et al. Antibiotic resistance is ancient. Nature. 2011;477(7365):457.

40. Ferreira LQ, Avelar KES, Vieira JMBD, de Paula GR, Colombo APV, Domingues RMCP, et al. Association Between the cfxA Gene and Transposon Tn4555 in Bacteroides distasonis Strains and Other Bacteroides Species. Current Microbiology. 2007;54(5):348-53.

41. Avelar KES, Otsuki K, Vicente ACP, Vieira JMBD, de Paula GR, Domingues RMCP, et al. Presence of the cfxA gene in Bacteroides distasonis. Research in Microbiology. 2003;154(5):369-74.

42. Salverda MLM, De Visser JAGM, Barlow M. Natural evolution of TEM-1 β-lactamase: experimental reconstruction and clinical relevance. FEMS Microbiology Reviews. 2010;34(6):1015-36.

43. Bengtsson-Palme J, Angelin M, Huss M, Kjellqvist S, Kristiansson E, Palmgren H, et al. The Human Gut Microbiome as a Transporter of Antibiotic Resistance Genes between Continents. Antimicrobial agents and chemotherapy. 2015;59(10):6551-60.

44. Doi Y, Park YS, Rivera JI, Adams-Haduch JM, Hingwe A, Sordillo EM, et al. Community-associated extended-spectrum β-lactamase-producing Escherichia coli infection in the United States. Clin Infect Dis. 2013;56(5):641-8.

45. Bag S, Ghosh TS, Banerjee S, Mehta O, Verma J, Dayal M, et al. Molecular insights into antimicrobial resistance traits of commensal human gut microbiota. Microbial ecology. 2019;77(2):546-57.

46. Jacoby GA, Strahilevitz J, Hooper DC. Plasmid-mediated quinolone resistance. Microbiology spectrum. 2014;2(5):10.1128/microbiolspec.PLAS-0006-2013.

47. von Wintersdorff CJH, Penders J, Stobberingh EE, Oude Lashof AML, Hoebe CJPA, Savelkoul PHM, et al. High rates of antimicrobial drug resistance gene acquisition after international travel, The Netherlands. Emerg Infect Dis. 2014;20(4):649-57.

48. Kim D-W, Thawng CN, Choi J-H, Lee K, Cha C-J. Polymorphism of antibiotic-inactivating enzyme driven by ecology expands the environmental resistome. ISME J. 2018;12(1):267-76.

49. Arzanlou M, Chai Wern C, Venter H. Intrinsic, adaptive and acquired antimicrobial resistance in Gram-negative bacteria. Essays in Biochemistry. 2017;61(1):49-59.

50. Fantin B, Duval X, Massias L, Alavoine L, Chau F, Retout S, et al. Ciprofloxacin dosage and emergence of resistance in human commensal bacteria. J Infect Dis. 2009;200(3):390-8.

51. Jacoby GA, Griffin CM, Hooper DC. Citrobacter spp. as a source of qnrB Alleles. Antimicrobial agents and chemotherapy. 2011;55(11):4979-84.

52. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proceedings of the National Academy of Sciences. 2011;108(supplement_1):4554-61.

53. Dethlefsen L, Huse S, Sogin ML, Relman DA. The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. PLOS Biology. 2008;6(11):e280.

54. Vollaard E, Clasener H. Colonization resistance. Antimicrobial agents and chemotherapy. 1994;38(3):409.

55. Baron SW, Ostrowsky BE, Nori P, Drory DY, Levi MH, Szymczak WA, et al. Screening of Clostridioides difficile carriers in an urban academic medical center: Understanding implications of disease. Infection Control & Hospital Epidemiology. 2020;41(2):149-53.

56. Crobach MJT, Vernon JJ, Loo VG, Kong LY, Péchiné S, Wilcox MH, et al. Understanding Clostridium difficile Colonization. Clin Microbiol Rev. 2018;31(2):e00021-17.

57. Brown KA, Khanafer N, Daneman N, Fisman DN. Meta-analysis of antibiotics and the risk of community-associated Clostridium difficile infection. Antimicrobial agents and chemotherapy. 2013;57(5):2326-32.

58. Theriot CM, Young VB. Interactions between the gastrointestinal microbiome and Clostridium difficile. Annual review of microbiology. 2015;69:445-61.

59. Nagaro KJ, Phillips ST, Cheknis AK, Sambol SP, Zukowski WE, Johnson S, et al. Nontoxigenic Clostridium difficile protects hamsters against challenge with historic and epidemic strains of toxigenic BI/NAP1/027 C. difficile. Antimicrobial agents and chemotherapy. 2013;57(11):5266-70.

60. Zhang L, Dong D, Jiang C, Li Z, Wang X, Peng Y. Insight into alteration of gut microbiota in Clostridium difficile infection and asymptomatic C. difficile colonization. Anaerobe. 2015;34:1-7.

61. Kamada N, Chen GY, Inohara N, Núñez G. Control of pathogens and pathobionts by the gut microbiota. Nat Immunol. 2013;14(7):685-90.

62. Mathers AJ, Peirano G, Pitout JDD. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae. Clin Microbiol Rev. 2015;28(3):565-91.

63. Österblad M, Hakanen A, Manninen R, Leistevuo T, Peltonen R, Meurman O, et al. A between-species comparison of antimicrobial resistance in enterobacteria in fecal flora. Antimicrobial Agents and Chemotherapy. 2000;44(6):1479-84.

64. Ravensbergen SJ, Berends M, Stienstra Y, Ott A. High prevalence of MRSA and ESBL among asylum seekers in the Netherlands. PLoS One. 2017;12(4).

65. Ravensbergen SJ, Louka C, Ott A, Rossen JW, Cornish D, Pournaras S, et al. Proportion of asylum seekers carrying multi-drug resistant microorganisms is persistently increased after arrival in the Netherlands. Antimicrobial Resistance & Infection Control. 2019;8(1):6.

66. Moreno E, Andreu A, Pigrau C, Kuskowski MA, Johnson JR, Prats G. Relationship between Escherichia coli strains causing acute cystitis in women and the fecal E. coli population of the host. J Clin Microbiol. 2008;46(8):2529-34.

67. Thänert R, Reske KA, Hink T, Wallace MA, Wang B, Schwartz DJ, et al. Comparative Genomics of Antibiotic-Resistant Uropathogens Implicates Three Routes for Recurrence of Urinary Tract Infections. mBio. 2019;10(4):e01977-19.

68. Moore AM, Patel S, Forsberg KJ, Wang B, Bentley G, Razia Y, et al. Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. PLoS One. 2013;8(11):e78822-e.

69. Mullany P. Functional metagenomics for the investigation of antibiotic resistance. Virulence. 2014;5(3):443-7.

70. Boolchandani M, Patel S, Dantas G. Functional Metagenomics to Study Antibiotic Resistance. In: Sass P, editor. Antibiotics: Methods and Protocols. New York, NY: Springer New York; 2017. p. 307-29.

71. Bertrand D, Shaw J, Kalathiyappan M, Ng AHQ, Kumar MS, Li C, et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. Nature Biotechnology. 2019;37(8):937-44.

72. Perron GG, Whyte L, Turnbaugh PJ, Goordial J, Hanage WP, Dantas G, et al. Functional Characterization of Bacteria Isolated from Ancient Arctic Soil Exposes Diverse Resistance Mechanisms to Modern Antibiotics. PLoS One. 2015;10(3):e0069533.

73. Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, et al. Bacterial phylogeny structures soil resistomes across habitats. Nature. 2014;509(7502):612-6.

74. Pehrsson EC, Tsukayama P, Patel S, Mejía-Bautista M, Sosa-Soto G, Navarrete KM, et al. Interconnected microbiomes and resistomes in low-income human habitats. Nature. 2016;533(7602):212-6.

75. Campbell TP, Sun X, Patel VH, Sanz C, Morgan D, Dantas G. The microbiome and resistome of chimpanzees, gorillas, and humans across host lifestyle and geography. ISME J. 2020.

76. Bae J, Lee KW, Islam MN, Yim H-S, Park H, Rho M. iMGEins: detecting novel mobile genetic elements inserted in individual genomes. BMC Genomics. 2018;19(1):944-.

77. Durrant MG, Li MM, Siranosian BA, Montgomery SB, Bhatt AS. A Bioinformatic Analysis of Integrative Mobile Genetic Elements Highlights Their Role in Bacterial Adaptation. Cell Host & Microbe. 2020;27(1):140-53.e9.

78. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics. 2015;13(5):278-89.

79. Meletis G, Chatzidimitriou D, Malisiovas N. Double- and multi-carbapenemase-producers: the excessively armored bacilli of the current decade. European Journal of Clinical Microbiology & Infectious Diseases. 2015;34(8):1487-93.

80. Huson DH, Albrecht B, Bağcı C, Bessarab I, Górska A, Jolic D, et al. MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. Biol Direct. 2018;13(1):6-.

81. Arango-Argoty GA, Dai D, Pruden A, Vikesland P, Heath LS, Zhang L. NanoARG: a web service for detecting and contextualizing antimicrobial resistance genes from nanopore-derived metagenomes. Microbiome. 2019;7(1):88.

82. Kintses B, Méhi O, Ari E, Számel M, Györkei Á, Jangir PK, et al. Phylogenetic barriers to horizontal transfer of antimicrobial peptide resistance genes in the human gut microbiota. Nature Microbiology. 2019;4(3):447-58.

83. Suneja G, Nain S, Sharma R. Microbiome: A Source of Novel Bioactive Compounds and Antimicrobial Peptides. In: Satyanarayana T, Johri BN, Das SK, editors. Microbial Diversity in Ecosystem Sustainability and Biotechnological Applications: Volume 1 Microbial Diversity in Normal & Extreme Environments. Singapore: Springer Singapore; 2019. p. 615-30.

84. Anthony WE, Burnham C-AD, Dantas G, Kwon JH. The Gut Microbiome as a Reservoir for Antimicrobial Resistance. J Infect Dis. 2020;223(Supplement_3):S209-S13.

85. Escherich T. The intestinal bacteria of the neonate and breast-fed infant. 1884. Rev Infect Dis. 1988;10(6):1220-5.

86. Nirenberg M, Leder P, Bernfield M, Brimacombe R, Trupin J, Rottman F, et al. RNA codewords and protein synthesis, VII. On the general nature of the RNA code. Proceedings of the National Academy of Sciences. 1965;53(5):1161-8.

87. Werkman CH. VITAMIN EFFECTS IN THE PHYSIOLOGY OF MICROORGANISMS. J Bacteriol. 1927;14(5):335-47.

88. Crick F, Barnett L, Brenner S, Watts-Tobin RJ. General nature of the genetic code for proteins. 1961.

89. Jiang W, Qiao JB, Bentley GJ, Liu D, Zhang F. Modular pathway engineering for the microbial production of branched-chain fatty alcohols. Biotechnology for Biofuels. 2017;10(1):244.

90. Wang C, Pfleger BF, Kim S-W. Reassessing Escherichia coli as a cell factory for biofuel production. Current Opinion in Biotechnology. 2017;45:92-103.

91. Kamionka M. Engineering of therapeutic proteins production in Escherichia coli. Curr Pharm Biotechnol. 2011;12(2):268-74.

92. Janßen HJ, Steinbüchel A. Fatty acid synthesis in Escherichia coli and its applications towards the production of fatty acid based biofuels. Biotechnology for Biofuels. 2014;7(1):7.

93. Hartl DL, Dykhuizen DE. The population genetics of Escherichia coli. Annual review of genetics. 1984;18(1):31-68.

94. Totsuka K. Studien über Bacterium coli. Zeitschrift für Hygiene und Infektionskrankheiten. 1903;45(1):115-24.

95. Blount ZD. The unexhausted potential of E. coli. eLife. 2015;4:e05826.

96. Bronfenbrenner JJ, Korb C. STUDIES ON THE BACTERIOPHAGE OF D'HÉRELLE: II. EFFECT OF ALCOHOL ON THE BACTERIOPHAGE OF D'HÉRELLE. The Journal of Experimental Medicine. 1925;42(3):419.

97. Cobb M. A breakthrough from 60 years ago: “General nature of the genetic code for proteins” (1961). Natural Sciences. 2021;1(1):e10018.

98. Linn S, Arber W. Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. Proceedings of the National Academy of Sciences. 1968;59(4):1300-6.

99. Cohen SN, Chang ACY, Boyer HW, Helling RB. Construction of Biologically Functional Bacterial Plasmids In Vitro. Proceedings of the National Academy of Sciences. 1973;70(11):3240-4.

100. Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of Escherichia coli K-12. Science. 1997;277(5331):1453-62.

101. Cooper TF. Recombination Speeds Adaptation by Reducing Competition between Beneficial Mutations in Populations of Escherichia coli. PLOS Biology. 2007;5(9):e225.

102. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, et al. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature. 2009;461(7268):1243-7.

103. Pruitt KD, Maglott DR. RefSeq and LocusLink: NCBI gene-centered resources. Nucleic acids research. 2001;29(1):137-40.

104. Keseler IM, Collado-Vides J, Gama-Castro S, Ingraham J, Paley S, Paulsen IT, et al. EcoCyc: a comprehensive database resource for Escherichia coli. Nucleic acids research. 2005;33(suppl_1):D334-D7.

105. Su C, Peregrin-Alvarez JM, Butland G, Phanse S, Fong V, Emili A, et al. Bacteriome. org—an integrated protein interaction database for E. coli. Nucleic acids research. 2007;36(suppl_1):D632-D6.

106. Guo AC, Jewison T, Wilson M, Liu Y, Knox C, Djoumbou Y, et al. ECMDB: the E. coli Metabolome Database. Nucleic acids research. 2012;41(D1):D625-D30.

107. Huerta AM, Salgado H, Thieffry D, Collado-Vides J. RegulonDB: a database on transcriptional regulation in Escherichia coli. Nucleic acids research. 1998;26(1):55-9.

108. Johnson IS. Human insulin from recombinant DNA technology. Science. 1983;219(4585):632-7.

109. Huang C-J, Lin H, Yang X. Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. Journal of Industrial Microbiology and Biotechnology. 2012;39(3):383-99.

110. Ponce E. Effect of growth rate reduction and genetic modifications on acetate accumulation and biomass yields in Escherichia coli. Journal of Bioscience and Bioengineering. 1999;87(6):775-80.

111. Aristidou AA, San K-Y, Bennett GN. Metabolic Engineering of Escherichia coli To Enhance Recombinant Protein Production through Acetate Reduction. Biotechnology Progress. 1995;11(4):475-8.

112. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):1-21.

113. Millard P, Enjalbert B, Uttenweiler-Joseph S, Portais J-C, Létisse F. Control and regulation of acetate overflow in Escherichia coli. eLife. 2021;10:e63661.

114. Bayer MH, Bayer ME. Phosphoglycerides and phospholipase C in membrane fractions of Escherichia coli B. Journal of Bacteriology. 1985;162(1):50-4.

115. Rowlett VW, Mallampalli VKPS, Karlstaedt A, Dowhan W, Taegtmeyer H, Margolin W, et al. Impact of Membrane Phospholipid Alterations in Escherichia coli on Cellular Function and Bacterial Stress Adaptation. Journal of Bacteriology. 2017;199(13).

116. Cronan JE, Jr. Phospholipid alterations during growth of Escherichia coli. Journal of bacteriology. 1968;95(6):2054-61.

117. Tang Y, Xia H, Li D. Membrane Phospholipid Biosynthesis in Bacteria. In: Cao Y, editor. Advances in Membrane Proteins: Part I: Mass Processing and Transportation. Singapore: Springer Singapore; 2018. p. 77-119.

118. Hildebrand JG, Law JH. Fatty Acid Distribution in Bacterial Phospholipids. The Specificity of the Cyclopropane Synthetase Reaction. Biochemistry. 1964;3:1304-8.

119. Chang Y-Y, Cronan JE. Membrane cyclopropane fatty acid content is a major factor in acid resistance of Escherichia coli. Molecular Microbiology. 1999;33(2):249-59.

120. Keweloh H, Diefenbach R, Rehm H-J. Increase of phenol tolerance of Escherichia coli by alterations of the fatty acid composition of the membrane lipids. Archives of Microbiology. 1991;157(1):49-53.

121. Wu T, Ye L, Zhao D, Li S, Li Q, Zhang B, et al. Membrane engineering - A novel strategy to enhance the production and accumulation of β-carotene in Escherichia coli. Metabolic Engineering. 2017;43:85-91.

122. Tan Z, Yoon JM, Nielsen DR, Shanks JV, Jarboe LR. Membrane engineering via trans unsaturated fatty acids production improves Escherichia coli robustness and production of biorenewables. Metabolic Engineering. 2016;35:105-13.

123. Lai CY, Cronan JE. Beta-ketoacyl-acyl carrier protein synthase III (FabH) is essential for bacterial fatty acid synthesis. J Biol Chem. 2003;278(51):51494-503.

124. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2:2006.0008.

125. Feng Y, Cronan JE. Complex binding of the FabR repressor of bacterial unsaturated fatty acid biosynthesis to its cognate promoters. Mol Microbiol. 2011;80(1):195-218.

126. Nunn WD, Giffin K, Clark D, Cronan JE, Jr. Role for fadR in unsaturated fatty acid biosynthesis in Escherichia coli. J Bacteriol. 1983;154(2):554-60.

127. Altabe SG, Aguilar P, Caballero GM, Mendoza DD. The Bacillus subtilis Acyl Lipid Desaturase Is a Δ5 Desaturase. Journal of Bacteriology. 2003;185(10):3228-31.

128. Grogan DW, Cronan JE. Cyclopropane ring formation in membrane lipids of bacteria. Microbiology and Molecular Biology Reviews. 1997;61(4):429-41.

129. Jovanovic S, Dietrich D, Becker J, Kohlstedt M, Wittmann C. Microbial production of polyunsaturated fatty acids — high-value ingredients for aquafeed, superfoods, and pharmaceuticals. Current Opinion in Biotechnology. 2021;69:199-211.

130. Hixson SM, Arts MT. Climate warming is predicted to reduce omega-3, long-chain, polyunsaturated fatty acid production in phytoplankton. Global Change Biology. 2016;22(8):2744-55.

131. Giner-Robles L, Lázaro B, de la Cruz F, Moncalián G. fabH deletion increases DHA production in Escherichia coli expressing Pfa genes. Microbial Cell Factories. 2018;17(1):88.

132. Maia MRG, Chaudhary LC, Bestwick CS, Richardson AJ, McKain N, Larson TR, et al. Toxicity of unsaturated fatty acids to the biohydrogenating ruminal bacterium, Butyrivibrio fibrisolvens. BMC microbiology. 2010;10(1):52.

133. Yazawa K. Production of eicosapentaenoic acid from marine bacteria. Lipids. 1996;31(1Part2):S297-S300.

134. Bonamore A, Macone A, Colotti G, Matarese RM, Boffi A. The desaturase from Bacillus subtilis, a promising tool for the selective olefination of phospholipids. Journal of Biotechnology. 2006;121(1):49-53.

135. Knothe G. Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. Fuel Processing Technology. 2005;86(10):1059-70.

136. Bentley GJ, Jiang W, Guamán LP, Xiao Y, Zhang F. Engineering Escherichia coli to produce branched-chain fatty acids in high percentages. Metabolic Engineering. 2016;38:148-58.

137. Machida S, Bakku RK, Suzuki I. Expression of Genes for a Flavin Adenine Dinucleotide-Binding Oxidoreductase and a Methyltransferase from Mycobacterium chlorophenolicum Is Necessary for Biosynthesis of 10-Methyl Stearic Acid from Oleic Acid in Escherichia coli. Frontiers in Microbiology. 2017;8.

138. Surger MJ, Angelov A, Stier P, Übelacker M, Liebl W. Impact of Branched-Chain Amino Acid Catabolism on Fatty Acid and Alkene Biosynthesis in Micrococcus luteus. Frontiers in Microbiology. 2018;9.

139. Health NIo. NIH guidelines for research involving recombinant DNA molecules (NIH guidelines). 2002.

140. Ahmad I, Nawaz N, Darwesh NM, ur Rahman S, Mustafa MZ, Khan SB, et al. Overcoming challenges for amplified expression of recombinant proteins using Escherichia coli. Protein Expression and Purification. 2018;144:12-8.

141. Huangfu J, Kim HS, Xu K, Ning X, Qin L, Li J, et al. Omics Analysis Reveals the Mechanism of Enhanced Recombinant Protein Production Under Simulated Microgravity. Frontiers in Bioengineering and Biotechnology. 2020;8.

142. Opgenorth P, Costello Z, Okada T, Goyal G, Chen Y, Gin J, et al. Lessons from Two Design–Build–Test–Learn Cycles of Dodecanol Production in Escherichia coli Aided by Machine Learning. ACS Synthetic Biology. 2019;8(6):1337-51.

143. Hodgman CE, Jewett MC. Cell-free synthetic biology: thinking outside the cell. Metab Eng. 2012;14(3):261-9.

144. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28(1):27-30.

145. Cho S, Choe D, Lee E, Kim SC, Palsson B, Cho B-K. High-Level dCas9 Expression Induces Abnormal Cell Morphology in Escherichia coli. ACS Synthetic Biology. 2018;7(4):1085-94.

146. Prasad PVV, Thomas JMG, Narayanan S. Global Warming Effects. In: Thomas B, Murray BG, Murphy DJ, editors. Encyclopedia of Applied Plant Sciences (Second Edition). Oxford: Academic Press; 2017. p. 289-99.

147. Zirogiannis N, Hollingsworth AJ, Konisky DM. Understanding Excess Emissions from Industrial Facilities: Evidence from Texas. Environmental Science & Technology. 2018;52(5):2482-90.

148. Lewis P, Karimi B, Shan Y, Rasdorf W. Comparing the economic, energy, and environmental impacts of biodiesel versus petroleum diesel fuel use in construction equipment. International Journal of Construction Education and Research. 2018:1-15.

149. Kalghatgi G. Is it really the end of internal combustion engines and petroleum in transport? Applied Energy. 2018;225:965-74.

150. Elgowainy A, Han J, Ward J, Joseck F, Gohlke D, Lindauer A, et al. Current and Future United States Light-Duty Vehicle Pathways: Cradle-to-Grave Lifecycle Greenhouse Gas Emissions and Economic Assessment. Environmental Science & Technology. 2018;52(4):2392-9.

151. ElGalad MI, El- Khatib KM, Abdelkader E, El-Araby R, ElDiwani G, Hawash SI. Empirical equations and economical study for blending biofuel with petroleum jet fuel. Journal of Advanced Research. 2018;9:43-50.

152. Coulman B, Dalai A, Heaton E, Lee CP, Lefsrud M, Levin D, et al. Developments in crops and management systems to improve lignocellulosic feedstock production. Biofuels, Bioproducts and Biorefining. 2013;7(5):582-601.

153. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. Lignin Biosynthesis and Structure. Plant Physiology. 2010;153(3):895-905.

154. Henson WR, Campbell T, DeLorenzo DM, Gao Y, Berla B, Kim SJ, et al. Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. Metabolic engineering. 2018;49:69-83.

155. Kurosawa K, Boccazzi P, de Almeida NM, Sinskey AJ. High-cell-density batch fermentation of Rhodococcus opacus PD630 using a high glucose concentration for triacylglycerol production. Journal of Biotechnology. 2010;147:212-8.

156. Alvarez HM, Mayer F, Fabritius D, Steinbüchel A. Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630. Archives of Microbiology. 1996;165(6):377-86.

157. EIA U. US crude oil and natural gas proved reserves, year-end 2015. US Energy Information Administration. 2016.

158. Bardi U. Peak oil, 20 years later: Failed prediction or useful insight? Energy Research & Social Science. 2019;48:257-61.

159. Pu Y, Zhang D, Singh PM, Ragauskas AJ. The new forestry biofuels sector. Biofuels, Bioproducts and Biorefining. 2008;2(1):58-73.

160. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, et al. The path forward for biofuels and biomaterials. Science. 2006;311(5760):484-9.

161. Vlachos DG, Chen JG, Gorte RJ, Huber GW, Tsapatsis M. Catalysis center for energy innovation for biomass processing: research strategies and goals. Catalysis letters. 2010;140(3-4):77-84.

162. Regalbuto JR. Cellulosic biofuels–got gasoline. Science. 2009;325(5942):822-4.

163. Kaparaju P, Serrano M, Thomsen AB, Kongjan P, Angelidaki I. Bioethanol, biohydrogen and biogas production from wheat straw in a biorefinery concept. Bioresource Technology. 2009;100(9):2562-8.

164. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, et al. Lignin valorization: improving lignin processing in the biorefinery. Science. 2014;344(6185):1246843.

165. Balan V, Chiaramonti D, Kumar S. Review of US and EU initiatives toward development, demonstration, and commercialization of lignocellulosic biofuels. Biofuels, Bioproducts and Biorefining. 2013;7(6):732-59.

166. Mottiar Y, Vanholme R, Boerjan W, Ralph J, Mansfield SD. Designer lignins: harnessing the plasticity of lignification. Current Opinion in Biotechnology. 2016;37:190-200.

167. Gou JY, Park S, Yu XH, Miller LM, Liu CJ. Compositional characterization and imaging of "wall-bound" acylesters of Populus trichocarpa reveal differential accumulation of acyl molecules in normal and reactive woods. Planta. 2008;229(1):15-24.

168. Boerjan W, Ralph J, Baucher M. Lignin biosynthesis. Annual review of plant biology. 2003;54(1):519-46.

169. Montague L, Slayton A, Lukas J. Lignocellulosic biomass to ethanol process design and economics utilizing co-current dilute acid prehydrolysis and enzymatic hydrolysis for corn stover. NREL Technical Report 2002; 2002.

170. Gao Y, Beganovic M, Foston MB. Lignin conversion to fuels and chemicals. Valorization of Lignocellulosic Biomass in a Biorefinery: From Logistics to Environmental and Performance Impact2016. p. 245-92.

171. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science. 2007;315(5813):804-7.

172. Babu BV. Biomass pyrolysis: a state-of-the-art review. Biofuels, Bioproducts and Biorefining. 2008;2(5):393-414.

173. Mohan D, Pittman, Charles U., Steele PH. Pyrolysis of Wood/Biomass for Bio-oil: A Critical Review. Energy & Fuels. 2006;20(3):848-89.

174. Xu P, Li L, Zhang F, Stephanopoulos G, Koffas M. Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(31):11299-304.

175. Zakzeski J, Bruijnincx PCA, Jongerius AL, Weckhuysen BM. The Catalytic Valorization of Lignin for the Production of Renewable Chemicals. Chemical Reviews. 2010;110(6):3552-99.

176. Zaheer M, Kempe R. Catalytic Hydrogenolysis of Aryl Ethers: A Key Step in Lignin Valorization to Valuable Chemicals. ACS Catalysis. 2015;5(3):1675-84.

177. Chen H, Liu J, Chang X, Chen D, Xue Y, Liu P, et al. A review on the pretreatment of lignocellulose for high-value chemicals. Fuel Processing Technology. 2017;160:196-206.

178. Bugg TD, Ahmad M, Hardiman EM, Rahmanpour R. Pathways for degradation of lignin in bacteria and fungi. Natural product reports. 2011;28(12):1883-96.

179. Li X, He Y, Zhang L, Xu Z, Ben H, Gaffrey MJ, et al. Discovery of potential pathways for biological conversion of poplar wood into lipids by co-fermentation of Rhodococci strains. Biotechnology for Biofuels. 2019;12(1):60.

180. Gasser CA, Hommes G, Schäffer A, Corvini PF-X. Multi-catalysis reactions: new prospects and challenges of biotechnology to valorize lignin. Applied Microbiology and Biotechnology. 2012;95(5):1115-34.

181. Beckham GT, Johnson CW, Karp EM, Salvachúa D, Vardon DR. Opportunities and challenges in biological lignin valorization. Current Opinion in Biotechnology. 2016;42:40-53.

182. Lynd LR, Van Zyl WH, McBride JE, Laser M. Consolidated bioprocessing of cellulosic biomass: an update. Current opinion in biotechnology. 2005;16(5):577-83.

183. Linger JG, Vardon DR, Guarnieri MT, Karp EM, Hunsinger GB, Franden MA, et al. Lignin valorization through integrated biological funneling and chemical catalysis. Proceedings of the National Academy of Sciences. 2014;111:12013-8.

184. Schwartz TJ, O'Neill BJ, Shanks BH, Dumesic JA. Bridging the Chemical and Biological Catalysis Gap: Challenges and Outlooks for Producing Sustainable Chemicals. ACS Catalysis. 2014;4:2060-9.

185. Sheldon RA. Green and sustainable manufacture of chemicals from biomass: state of the art. Green Chemistry. 2014;16:950-63.

186. Taherzadeh MJ, Karimi K. ACID-BASED HYDROLYSIS PROCESSES FOR ETHANOL FROM LIGNOCELLULOSIC MATERIALS: A REVIEW2007.

187. Rover MR, Johnston PA, Jin T, Smith RG, Brown RC, Jarboe L. Production of Clean Pyrolytic Sugars for Fermentation. ChemSusChem. 2014;7(6):1662-8.

188. Okamura-Abe Y, Abe T, Nishimura K, Kawata Y, Sato-Izawa K, Otsuka Y, et al. Beta-ketoadipic acid and muconolactone production from a lignin-related aromatic compound through the protocatechuate 3,4-metabolic pathway. Journal of Bioscience and Bioengineering. 2016;121:652-8.

189. Mycroft Z, Gomis M, Mines P, Law P, Bugg TDH. Biocatalytic conversion of lignin to aromatic dicarboxylic acids in Rhodococcus jostii RHA1 by re-routing aromatic degradation pathways. Green Chemistry. 2015;17:4974-9.

190. Matera I, Ferraroni M, Kolomytseva M, Golovleva L, Scozzafava A, Briganti F. Catechol 1,2-dioxygenase from the Gram-positive Rhodococcus opacus 1CP: Quantitative structure/activity relationship and the crystal structures of native enzyme and catechols adducts. Journal of Structural Biology. 2010;170:548-64.

191. Weber C, Brückner C, Weinreb S, Lehr C, Essl C, Boles E. Biosynthesis of cis,cis-Muconic Acid and Its Aromatic Precursors, Catechol and Protocatechuic Acid, from Renewable Feedstocks by Saccharomyces cerevisiae. Applied and Environmental Microbiology. 2012;78:8421-30.

192. Sonoki T, Morooka M, Sakamoto K, Otsuka Y, Nakamura M, Jellison J, et al. Enhancement of protocatechuate decarboxylase activity for the effective production of muconate from lignin-related aromatic compounds. Journal of Biotechnology. 2014;192 Pt A:71-7.

193. Wei Z, Zeng G, Huang F, Kosa M, Sun Q, Meng X, et al. Microbial lipid production by oleaginous Rhodococci cultured in lignocellulosic autohydrolysates. Applied microbiology and biotechnology. 2015;99(17):7369-77.

194. Wells T, Wei Z, Ragauskas A. Bioconversion of lignocellulosic pretreatment effluent via oleaginous Rhodococcus opacus DSM 1069. Biomass and Bioenergy. 2015;72:200-5.

195. Linger JG, Vardon DR, Guarnieri MT, Karp EM, Hunsinger GB, Franden MA, et al. Lignin valorization through integrated biological funneling and chemical catalysis. Proceedings of the National Academy of Sciences. 2014;111(33):12013-8.

196. Wei Z, Zeng G, Huang F, Kosa M, Huang D, Ragauskas AJ. Bioconversion of oxygen-pretreated Kraft lignin to microbial lipid with oleaginous Rhodococcus opacus DSM 1069. Green Chemistry. 2015;17(5):2784-9.

197. Wei Z, Zeng G, Kosa M, Huang D, Ragauskas AJ. Pyrolysis oil-based lipid production as biodiesel feedstock by Rhodococcus opacus. Applied biochemistry and biotechnology. 2015;175(2):1234-46.

198. Kurosawa K, Laser J, Sinskey AJ. Tolerance and adaptive evolution of triacylglycerol-producing Rhodococcus opacus to lignocellulose-derived inhibitors. Biotechnology for Biofuels. 2015;8(1):76.

199. Yoneda A, Henson WR, Goldner NK, Park KJ, Forsberg KJ, Kim SJ, et al. Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. Nucleic Acids Res. 2016;44(5):2240-54.

200. Wang B, Rezenom YH, Cho K-C, Tran JL, Lee DG, Russell DH, et al. Cultivation of lipid-producing bacteria with lignocellulosic biomass: Effects of inhibitory compounds of lignocellulosic hydrolysates. Bioresource Technology. 2014;161:162-70.

201. Yang S, Franden MA, Yang Q, Chou Y-C, Zhang M, Pienkos PT. Identification of Inhibitors in Lignocellulosic Slurries and Determination of Their Effect on Hydrocarbon-Producing Microorganisms. Frontiers in bioengineering and biotechnology. 2018;6:23-.

202. Xie S, Sun Q, Pu Y, Lin F, Sun S, Wang X, et al. Advanced Chemical Design for Efficient Lignin Bioconversion. ACS Sustainable Chemistry & Engineering. 2017;5(3):2215-23.

203. Holder JW, Ulrich JC, DeBono AC, Godfrey PA, Desjardins CA, Zucker J, et al. Comparative and Functional Genomics of Rhodococcus opacus PD630 for Biofuels Development. PLOS Genetics. 2011;7(9):e1002219.

204. Zhao C, Xie S, Pu Y, Zhang R, Huang F, Ragauskas AJ, et al. Synergistic enzymatic and microbial lignin conversion. Green Chemistry. 2016;18(5):1306-12.

205. Le RK, Wells Jr T, Das P, Meng X, Stoklosa RJ, Bhalla A, et al. Conversion of corn stover alkaline pre-treatment waste streams into biodiesel via Rhodococci. RSC Advances. 2017;7(7):4108-15.

206. He Y, Li X, Ben H, Xue X, Yang B. Lipid Production from Dilute Alkali Corn Stover Lignin by Rhodococcus Strains. ACS Sustainable Chemistry & Engineering. 2017;5(3):2302-11.

207. Xiong X, Wang X, Chen S. Engineering of a xylose metabolic pathway in Rhodococcus strains. Applied and environmental microbiology. 2012;78(16):5483-91.

208. Hetzler S, Bröker D, Steinbüchel A. Saccharification of Cellulose by Recombinant Rhodococcus opacus PD630 Strains. Applied and environmental microbiology. 2013;79(17):5159-66.

209. Kurosawa K, Plassmeier J, Kalinowski J, Rückert C, Sinskey AJ. Engineering l-arabinose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production. Metabolic Engineering. 2015;30:89-95.

210. Chen Y, Ding Y, Yang L, Yu J, Liu G, Wang X, et al. Integrated omics study delineates the dynamics of lipid droplets in Rhodococcus opacus PD630. Nucleic Acids Res. 2014;42(2):1052-64.

211. Alvarez AF, Alvarez HM, Kalscheuer R, Wältermann M, Steinbüchel A. Cloning and characterization of a gene involved in triacylglycerol biosynthesis and identification of additional homologous genes in the oleaginous bacterium Rhodococcus opacus PD630. 2008;154(8):2327-35.

212. Tsitko IV, Zaitsev GM, Lobanok AG, Salkinoja-Salonen MS. Effect of Aromatic Compounds on Cellular Fatty Acid Composition of Rhodococcus opacus. Applied and Environmental Microbiology. 1999;65(2):853-5.

213. Kumar A, Kumar S, Kumar S. Biodegradation kinetics of phenol and catechol using Pseudomonas putida MTCC 1194. Biochemical Engineering Journal. 2005;22(2):151-9.

214. Arutchelvan V, Kanakasabai V, Elangovan R, Nagarajan S, Muralikrishnan V. Kinetics of high strength phenol degradation using Bacillus brevis. Journal of Hazardous Materials. 2006;129(1):216-22.

215. Mahan KM, Le RK, Yuan J, Ragauskas AJ. A Review on The Bioconversion of Lignin to Microbial Lipid with Oleaginous. Journal of Biotechnology & Biomaterials. 2017;7(2):1.

216. Hollinshead WD, Henson WR, Abernathy M, Moon TS, Tang YJ. Rapid metabolic analysis of Rhodococcus opacus PD630 via parallel 13C-metabolite fingerprinting. Biotechnology and Bioengineering. 2016;113(1):91-100.

217. Ravi K, Abdelaziz OY, Nöbel M, García-Hidalgo J, Gorwa-Grauslund MF, Hulteberg CP, et al. Bacterial conversion of depolymerized Kraft lignin. Biotechnology for Biofuels. 2019;12(1):56.

218. DeLorenzo DM, Rottinghaus AG, Henson WR, Moon TS. Molecular Toolkit for Gene Expression Control and Genome Modification in Rhodococcus opacus PD630. ACS Synthetic Biology. 2018;7(2):727-38.

219. Eulberg D, Schlomann M. The putative regulator of catechol catabolism in Rhodococcus opacus 1CP--an IclR-type, not a LysR-type transcriptional regulator. Antonie van Leeuwenhoek. 1998;74(1-3):71-82.

220. Di Canito A, Zampolli J, Orro A, D'Ursi P, Milanesi L, Sello G, et al. Genome-based analysis for the identification of genes involved in o-xylene degradation in Rhodococcus opacus R7. BMC Genomics. 2018;19(1):587.

221. Kitagawa W, Kimura N, Kamagata Y. A Novel p-Nitrophenol Degradation Gene Cluster from a Gram-Positive Bacterium, Rhodococcus opacus SAO101. Journal of Bacteriology. 2004;186(15):4894-902.

222. Martinkova L, Uhnakova B, Patek M, Nesvera J, Kren V. Biodegradation potential of the genus Rhodococcus. Environment international. 2009;35(1):162-77.

223. Zaitsev GM, Uotila JS, Tsitko IV, Lobanok AG, Salkinoja-Salonen MS. Utilization of Halogenated Benzenes, Phenols, and Benzoates by Rhodococcus opacus GM-14. Appl Environ Microbiol. 1995;61(12):4191-201.

224. van Kessel JC, Hatfull GF. Recombineering in Mycobacterium tuberculosis. Nature methods. 2007;4(2):147-52.

225. Larkin MJ, Kulakov LA, Allen CCR. Biodegradation and Rhodococcus – masters of catabolic versatility. Current Opinion in Biotechnology. 2005;16(3):282-90.

226. Ehrt S, Guo XV, Hickey CM, Ryou M, Monteleone M, Riley LW, et al. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. Nucleic Acids Res. 2005;33(2):e21.

227. Kurosawa K, Wewetzer SJ, Sinskey AJ. Engineering xylose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production. Biotechnol Biofuels. 2013;6(1):134.

228. Pátek M, Nešvera J. Promoters and Plasmid Vectors of Corynebacterium glutamicum. In: Yukawa H, Inui M, editors. Corynebacterium glutamicum. Microbiology Monographs: Springer Berlin Heidelberg; 2013. p. 51-88.

229. Zhang Y, Praszkier J, Hodgson A, Pittard AJ. Molecular analysis and characterization of a broad-host-range plasmid, pEP2. J Bacteriol. 1994;176(18):5718-28.

230. Vesely M, Patek M, Nesvera J, Cejkova A, Masak J, Jirku V. Host-vector system for phenol-degrading Rhodococcus erythropolis based on Corynebacterium plasmids. Appl Microbiol Biotechnol. 2003;61(5-6):523-7.

231. Archer JA, Sinskey AJ. The DNA sequence and minimal replicon of the Corynebacterium glutamicum plasmid pSR1: evidence of a common ancestry with plasmids from C. diphtheriae. Journal of general microbiology. 1993;139(8):1753-9.

232. Nesvera J, Patek M, Hochmannova J, Abrhamova Z, Becvarova V, Jelinkova M, et al. Plasmid pGA1 from Corynebacterium glutamicum codes for a gene product that positively influences plasmid copy number. J Bacteriol. 1997;179(5):1525-32.

233. Ellinger J, Schmidt-Dannert C. Construction of a BioBrick compatible vector system for Rhodococcus. Plasmid. 2017;90:1-4.

234. Lessard PA, O'Brien XM, Currie DH, Sinskey AJ. pB264, a small, mobilizable, temperature sensitive plasmid from Rhodococcus. BMC microbiology. 2004;4:15.

235. Ranes MG, Rauzier J, Lagranderie M, Gheorghiu M, Gicquel B. Functional analysis of pAL5000, a plasmid from Mycobacterium fortuitum: construction of a "mini" mycobacterium-Escherichia coli shuttle vector. J Bacteriol. 1990;172(5):2793-7.

236. Mahenthiralingam E, Draper P, Davis EO, Colston MJ. Cloning and sequencing of the gene which encodes the highly inducible acetamidase of Mycobacterium smegmatis. Journal of general microbiology. 1993;139(3):575-83.

237. Timm J, Lim EM, Gicquel B. Escherichia coli-mycobacteria shuttle vectors for operon and gene fusions to lacZ: the pJEM series. J Bacteriol. 1994;176(21):6749-53.

238. Snapper SB, Melton RE, Mustafa S, Kieser T, Jr WRJ. Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. Molecular Microbiology. 1990;4(11):1911-9.

239. Chiba K, Hoshino Y, Ishino K, Kogure T, Mikami Y, Uehara Y, et al. Construction of a pair of practical Nocardia-Escherichia coli shuttle vectors. Japanese Journal of Infectious Diseases. 2007;60(1):45-7.

240. Tauch A, Kirchner O, Loffler B, Gotker S, Puhler A, Kalinowski J. Efficient electrotransformation of corynebacterium diphtheriae with a mini-replicon derived from the Corynebacterium glutamicum plasmid pGA1. Current microbiology. 2002;45(5):362-7.

241. Hetzler S, Steinbüchel A. Establishment of Cellobiose Utilization for Lipid Production in Rhodococcus opacus PD630. Applied and Environmental Microbiology. 2013;79(9):3122-5.

242. Kalscheuer R, Arenskotter M, Steinbuchel A. Establishment of a gene transfer system for Rhodococcus opacus PD630 based on electroporation and its application for recombinant biosynthesis of poly(3-hydroxyalkanoic acids). Appl Microbiol Biotechnol. 1999;52(4):508-15.

243. Hernandez MA, Arabolaza A, Rodriguez E, Gramajo H, Alvarez HM. The atf2 gene is involved in triacylglycerol biosynthesis and accumulation in the oleaginous Rhodococcus opacus PD630. Appl Microbiol Biotechnol. 2013;97(5):2119-30.

244. Lanfranconi MP, Alvarez HM. Rewiring neutral lipids production for the de novo synthesis of wax esters in Rhodococcus opacus PD630. Journal of Biotechnology. 2017;260:67-73.

245. Nga DP, Altenbuchner J, Heiss GS. NpdR, a repressor involved in 2,4,6-trinitrophenol degradation in Rhodococcus opacus HL PM-1. J Bacteriol. 2004;186(1):98-103.

246. Pelicic V, Reyrat JM, Gicquel B. Expression of the Bacillus subtilis sacB gene confers sucrose sensitivity on mycobacteria. J Bacteriol. 1996;178(4):1197-9.

247. Nakashima N, Tamura T. A novel system for expressing recombinant proteins over a wide temperature range from 4 to 35°C. Biotechnology and Bioengineering. 2004;86(2):136-48.

248. Dong L, Nakashima N, Tamura N, Tamura T. Isolation and characterization of the Rhodococcus opacus thiostrepton-inducible genes tipAL and tipAS: application for recombinant protein expression in Rhodococcus. FEMS microbiology letters. 2004;237(1):35-40.

249. DeLorenzo DM, Henson WR, Moon TS. Development of Chemical and Metabolite Sensors for Rhodococcus opacus PD630. ACS Synthetic Biology. 2017;6(10):1973-8.

250. Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD, Rehrauer WM. Biochemistry of homologous recombination in Escherichia coli. Microbiological Reviews. 1994;58(3):401-65.

251. Furuya N, Komano T. Initiation and termination of DNA transfer during conjugation of IncI1 plasmid R64: roles of two sets of inverted repeat sequences within oriT in termination of R64 transfer. J Bacteriol. 2000;182(11):3191-6.

252. Sugar IP, Neumann E. Stochastic model for electric field-induced membrane pores electroporation. Biophysical Chemistry. 1984;19(3):211-25.

253. Gilman J, Love J. Synthetic promoter design for new microbial chassis. Biochemical Society Transactions. 2016;44(3):731-7.

254. Coussement P, Bauwens D, Maertens J, De Mey M. Direct Combinatorial Pathway Optimization. ACS Synthetic Biology. 2017;6(2):224-32.

255. Desomer J, Crespi M, Van Montagu M. Illegitimate integration of non-replicative vectors in the genome of Rhodococcus fascians upon electrotransformation as an insertional mutagenesis system. Mol Microbiol. 1991;5(9):2115-24.

256. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nature protocols. 2013;8(11):2180-96.

257. Rock JM, Hopkins FF, Chavez A, Diallo M, Chase MR, Gerrick ER, et al. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. Nat Microbiol. 2017;2:16274.

258. DeLorenzo DM, Moon TS. Selection of stable reference genes for RT-qPCR in Rhodococcus opacus PD630. Sci Rep. 2018;8(1):6019.

259. Gill JJ, Wang B, Sestak E, Young R, Chu K-H. Characterization of a Novel Tectivirus Phage Toil and Its Potential as an Agent for Biolipid Extraction. Scientific Reports. 2018;8(1):1062.

260. Kurosawa K, Wewetzer SJ, Sinskey AJ. Triacylglycerol Production from Corn Stover Using a Xylose-Fermenting Rhodococcus opacus Strain for Lignocellulosic Biofuels. Journal of Microbial & Biochemical Technology. 2014;6(5):254-9.

261. Suwaleerat T, Thanapimmetha A, Srisaiyoot M, Chisti Y, Srinophakun P. Enhanced production of carotenoids and lipids by Rhodococcus opacus PD630. Journal of Chemical Technology & Biotechnology. 2018;93(8):2160-9.

262. Herrero OM, Moncalián G, Alvarez HM. Physiological and genetic differences amongst Rhodococcus species for using glycerol as a source for growth and triacylglycerol production. Microbiology. 2016;162(2):384-97.

263. Goswami L, Tejas Namboodiri MM, Vinoth Kumar R, Pakshirajan K, Pugazhenthi G. Biodiesel production potential of oleaginous Rhodococcus opacus grown on biomass gasification wastewater. Renewable Energy. 2017;105:400-6.

264. Kurosawa K, Radek A, Plassmeier JK, Sinskey AJ. Improved glycerol utilization by a triacylglycerol-producing Rhodococcus opacus strain for renewable fuels. Biotechnol Biofuels. 2015;8:31.

265. Huang L, Zhao L, Zan X, Song Y, Ratledge C. Boosting fatty acid synthesis in Rhodococcus opacus PD630 by overexpression of autologous thioesterases. Biotechnology Letters. 2016;38(6):999-1008.

266. Herrero OM, Alvarez HM. Whey as a renewable source for lipid production by Rhodococcus strains: Physiology and genomics of lactose and galactose utilization. European Journal of Lipid Science and Technology. 2016;118(2):262-72.

267. Herrero OM, Villalba MS, Lanfranconi MP, Alvarez HM. Rhodococcus bacteria as a promising source of oils from olive mill wastes. World Journal of Microbiology and Biotechnology. 2018;34(8):114.

268. Sainsbury PD, Hardiman EM, Ahmad M, Otani H, Seghezzi N, Eltis LD, et al. Breaking down lignin to high-value chemicals: the conversion of lignocellulose to vanillin in a gene deletion mutant of Rhodococcus jostii RHA1. ACS chemical biology. 2013;8(10):2151-6.

269. Salvachúa D, Johnson CW, Singer CA, Rohrer H, Peterson DJ, Black BA, et al. Bioprocess development for muconic acid production from aromatic compounds and lignin. Green Chemistry. 2018;20(21):5007-19.

270. Becker J, Kuhl M, Kohlstedt M, Starck S, Wittmann C. Metabolic engineering of Corynebacterium glutamicum for the production of cis, cis-muconic acid from lignin. Microbial cell factories. 2018;17(1):115.

271. Shi Y, Yan X, Li Q, Wang X, Xie S, Chai L, et al. Directed bioconversion of Kraft lignin to polyhydroxyalkanoate by Cupriavidus basilensis B-8 without any pretreatment. Process Biochemistry. 2017;52:238-42.

272. Bradfield MF, Mohagheghi A, Salvachúa D, Smith H, Black BA, Dowe N, et al. Continuous succinic acid production by Actinobacillus succinogenes on xylose-enriched hydrolysate. Biotechnology for biofuels. 2015;8(1):181.

273. Anthony WE, Carr RR, DeLorenzo DM, Campbell TP, Shang Z, Foston M, et al. Development of Rhodococcus opacus as a chassis for lignin valorization and bioproduction of high-value compounds. Biotechnology for Biofuels. 2019;12(1):192.

Chapter 2:

Acute and persistent effects of commonly-used antibiotics on the gut microbiome and resistome in healthy adults

2.1 Introduction

2.1.1 Abstract

Antibiotics are deployed against bacterial pathogens, but their targeting of conserved microbial processes means they also collaterally perturb the commensal microbiome. To understand acute and persistent impacts of antibiotics on the gut microbiota of healthy adult volunteers, we quantify microbiome dynamics before, during, and 6 months after exposure to 4 commonly used antibiotic regimens. We observe an acute decrease in species richness and culturable bacteria after antibiotics, with most healthy adult microbiomes returning to pre-treatment species richness after 2 months, but with an altered taxonomy, resistome, and metabolic output, as well as an increased antibiotic resistance burden. Azithromycin delays recovery of species richness, resulting in greater compositional distance. A subset of volunteers experience a persistent reduction in microbiome diversity after antibiotics and share compositional similarities with patients hospitalized in intensive care units. These results improve our quantitative understanding of the impact of antibiotics on commensal microbiome dynamics, resilience, and recovery.

2.1.2 Introduction

Antibiotics are critical, life-saving medications that have served as the primary treatment for bacterial infections since their clinical introduction in the 1940s (1, 2). However, antibiotic exposure can also result in significant unintended consequences, including acute and persistent changes in the commensal host microbiome (3-7), as well as selection for antibiotic resistance (AR) genes in commensals and pathogens (8, 9). Immediately after antibiotic usage, changes in the microbiome can leave patients more susceptible to re-infection or infection from opportunistic pathogens, such as *Clostridioides difficile* (10, 11). Further, taxonomic alterations to the microbiome due to antibiotics can lead to long-term consequences such as slower development of a diverse microbiome in pre-term infants (8, 9, 12). Such *gut microbiome dysbioses*, commonly observed with broad-spectrum antibiotic treatment in patients hospitalized in intensive care units (ICU) (13, 14), have been linked to hospital acquired infections (15) and multiple organ dysfunction syndrome (16). Resistome selection and enrichment can also persist over long periods, as shown in studies where pre-term infants (8, 9, 12), nursing home residents (17), and the elderly (18, 19) have increased AR organism and AR gene carriage, which has been theorized to be caused in part by repeated exposure to antibiotic treatment.

Existing literature explaining the effects of clinically relevant antibiotic exposure on the taxonomic and functional architecture of the healthy human microbiome is limited, as healthy individuals do not routinely need antibiotics. Two foundational studies assessed the effect of a short course of ciprofloxacin on the taxonomic structure of the microbiomes of 3 healthy individuals, using pyrosequencing of the 16S rRNA gene hypervariable region (4, 20). The authors observed an almost universal acute decrease in taxonomic diversity after treatment and an incomplete recovery in some cases, but they were unable to assess changes in AR genes or

other functional components of the microbiome because of the methodologies employed. Similar findings have been reported for amoxicillin (21) and the amoxicillin-clavulanic acid combination (22), indicative of an interplay between more universal signatures of bacterial selection and person-to-person variability of outcomes after antibiotic treatment. A few recent studies have applied shotgun metagenomics to this question, which can assess not only variation in taxonomic diversity but also resistome abundance and composition. These studies demonstrated an effect of moxifloxacin on bacterial gene richness within the first 5 weeks (7), a significant increase in *tet* resistance genes after 1 week in healthy volunteers given amoxicillin (23), and a partial recovery of species richness after a 4-day course of a 3-drug cocktail in a cohort of 18 healthy young men (6). While these studies have increased our understanding of taxonomic and resistance composition changes beyond what 16S rRNA sequencing can describe, they are still limited by the number of treatments and timepoints studied, and thus a cohesive understanding of both the acute and chronic effects for commonly prescribed antibiotics is still missing. The concept of ecological opportunity is used to describe environments with niche availability and discordance that drive persistence and adaption in community members (24, 25). Gut microbiomes have been theorized to be environments with great ecological opportunity (26, 27), especially during low diversity events such as infant development (28, 29), and perturbation due to antibiotics or disease (26). There are open questions concerning the effect of time-induced drift during microbiome recovery: as the community begins the process of reassembly in the presence of external factors (antibiotics), do treatments vary in the time-to-recovery and net compositional differences? What genetic components and commensal bacteria define the ecological succession?

A complementary approach to prospectively investigate antibiotic perturbation of the microbiome is to employ laboratory murine models, where many of the confounders from a

retrospective human cohort can be controlled for. Murine models have shown antibiotic perturbation alongside diet can modulate compositional differences in the microbiome (30, 31) and host susceptibility to infection by *Salmonella enterica* (32) and *C. difficile* (33, 34). While these murine studies have served important roles in hypothesis testing and mechanistic exploration (35, 36), it is often difficult to translate these findings to human health and treatment because of the substantial differences in 'natural' human and mouse microbiomes and host biology (37), and the idealized nature of experimental work in mice (inbred mouse lines often fed restricted, highly regulated diets) (38).

We hypothesized that antibiotic regimens which are routinely used for treatment of bacterial infections would cause acute perturbation to the microbiomes of healthy human volunteers by increasing the relative abundance of pathobionts and select for AR genes. We further hypothesized that while the taxonomic diversity of these microbiomes would largely recover to pre-perturbation states within weeks to a few months, the 'scars' of perturbed taxonomic composition, AR gene enrichment, and compositional drift, would persist over the same time scales. Thus, in this study we designed and executed a prospective, longitudinal investigation of the impact of four different antibiotic regimens recommended for outpatient community associated pneumonia on the taxonomy, resistome, and functional output of healthy volunteer microbiomes. This represents a patient population who would receive antibiotics for respiratory infections which are frequently not bacterial in etiology. Thus, many antibiotics are frequently misused for that indication, and we were accordingly motivated to understand the effect of these common unwarranted treatments on a healthy human microbiome. In our study design, we collected multiple control samples per individual before antibiotic exposure to establish pre-perturbation baseline microbiome states. Using both quantitative microbiologic

culture and metagenomic sequencing, we analyzed both the acute effect of antibiotic perturbation (AP) on healthy volunteer gut microbiomes and the putative persistent "antibiotic scarring" (defined as a statistically significant increase in AR compared to the pre-antibiotic state) observed up to 6 months after antibiotic exposure, and we compared the trajectories of these perturbations to the microbiomes of hospitalized, critically-ill patients.

We found a remarkable similarity in the effect of each antibiotic treatment on species richness and viable colony forming units (CFU) concentrations with antibiotic specific enrichment in taxonomy and resistance. Recovery after treatments containing azithromycin was slower, resulting in niche discordance and greater net compositional distance. Three volunteers already at low microbiome diversity seemed primed to large taxonomic perturbations after antibiotic treatment, increasing their taxonomic similarity to microbiomes from critically-ill ICU patients. Even though most volunteers returned to a similar level of taxonomic diversity after treatment, significant increases in AR genes, changes to the functional output of the microbiome, and an altered taxonomic state imply significant long-term consequences. These data quantitatively illuminate how the healthy microbiome buffers antibiotic specific changes to the relative abundance of specific taxa during perturbation, and identifies key areas of further research, such as identifying individuals at greater risk of perturbation.

The bulk of this research on the impact of antibiotics on the human microbiome and resistome has been understandably performed with retrospective cohorts of severely-ill and hospitalized individuals, who are at high risk for infections and who accordingly receive many courses of prophylactic and empiric antibiotic therapy (13, 14, 39). Antibiotic exposure in these populations occurs in the context of diverse confounding factors, such as infection (13, 14), drug exposure (12), hospital environment (8, 14), and potential immunocompromise (12). Statistical

modeling, which allows for some control of these confounders, demonstrates that the state of the microbiome pre-antibiotic exposure has some predictive power on the impact and severity of antibiotic exposure perturbation on the microbiome, but these models are still limited by the immense variation observed between individuals studied retrospectively (8, 40). To separate the acute and persistent impacts on microbiome composition and function from antibiotic exposure versus from the effects of illness and hospitalization, we must look to prospective studies that characterize the impact of antibiotics on healthy, un-hospitalized adults (4, 6, 7, 20). Further, the healthy, adult microbiome is relatively stable in an individual and has been theorized to be robust and resilient to perturbation (41). Yet adult microbiomes can vary considerably across individuals (42, 43), and this variability may result in different levels of vulnerability to and severity of AR perturbation. Understanding these differences and defining the effects of specific antibiotics on healthy human microbiomes can help identify predictive factors that distinguish healthy and dysbiotic states, informing development of future anti-infective treatments which minimize collateral microbiome damage (40).

2.2 Results

To quantify the effect of antibiotics on the gut microbiota of healthy individuals over time, we recruited a study group of 20 volunteers from the St. Louis, MO, metropolitan area (Table S1). The cohort was then randomized into one of four antibiotic treatment groups: 1. azithromycin (AZM) dispensed as a standard oral dose pack with a 500mg first dose on day 1 and 250mg daily thereafter, 2. levofloxacin (LVX) dispensed as 750mg, 3. cefpodoxime (CPD), dispensed as 250mg twice a day, and 4. a combination of the azithromycin and cefpodoxime (CPD+AZM) treatments at the aforementioned individual doses. All treatments were administered orally for 5 days, consistent with Infectious Diseases Society of America/American

Thoracic Society guidelines for community acquired pneumonia in adults (44). Fecal samples from volunteers were collected longitudinally at 15 timepoints encompassing periods before, during, and after antibiotic treatment (Table S2). The first 4 sampling points were taken before antibiotic administration, to establish a robust baseline for each patient's unperturbed microbiome over a timespan similar to the period of acute perturbation after treatment (2 weeks). Antibiotic treatments started on day 0, and samples were taken on days 3, 6, 9, 12, and 19 post start of antibiotics. After the first 30 days post-treatment, sampling intervals were increased to a month (days 35, 65, 95, 125), and finally 2 months (day 185) for the last interval. The participant retention rate was 100%, with a 96.3% sample submission rate, for a total of 289 fecal samples for analysis. A subset of 10 sampling points per individual underwent semi-quantitative microbiologic culture. The microbiomes of this healthy volunteer cohort were then compared to fecal samples from 26 ICU patients in the St. Louis, MO, area who were being screened for C. difficile colonization. ICU patients are a patient population with a well-documented high AR burden (45, 46).

2.2.1 Antibiotics decrease microbiome bacterial load and richness and perturb microbial community structure

We analyzed gut microbiomes using complementary culture-based and culture-independent methods. To assess the effect of antibiotics on the bacterial load of the gut microbiome, we used semi-quantitative microbiologic culture of viable bacteria in a subset of fecal samples. We found a significant mean reduction of 4.78 log-transformed CFU (CI [2.89–6.36], p = 0.0074 [paired Wilcoxon rank sum]) and 2.90 (CI [1.90–3.91], p = 0.0047) in aerobic and anaerobic bacterial titers, respectively, between day 6 and day -14 (Figure 2.1A-B, Table 2.3S-2.4S).

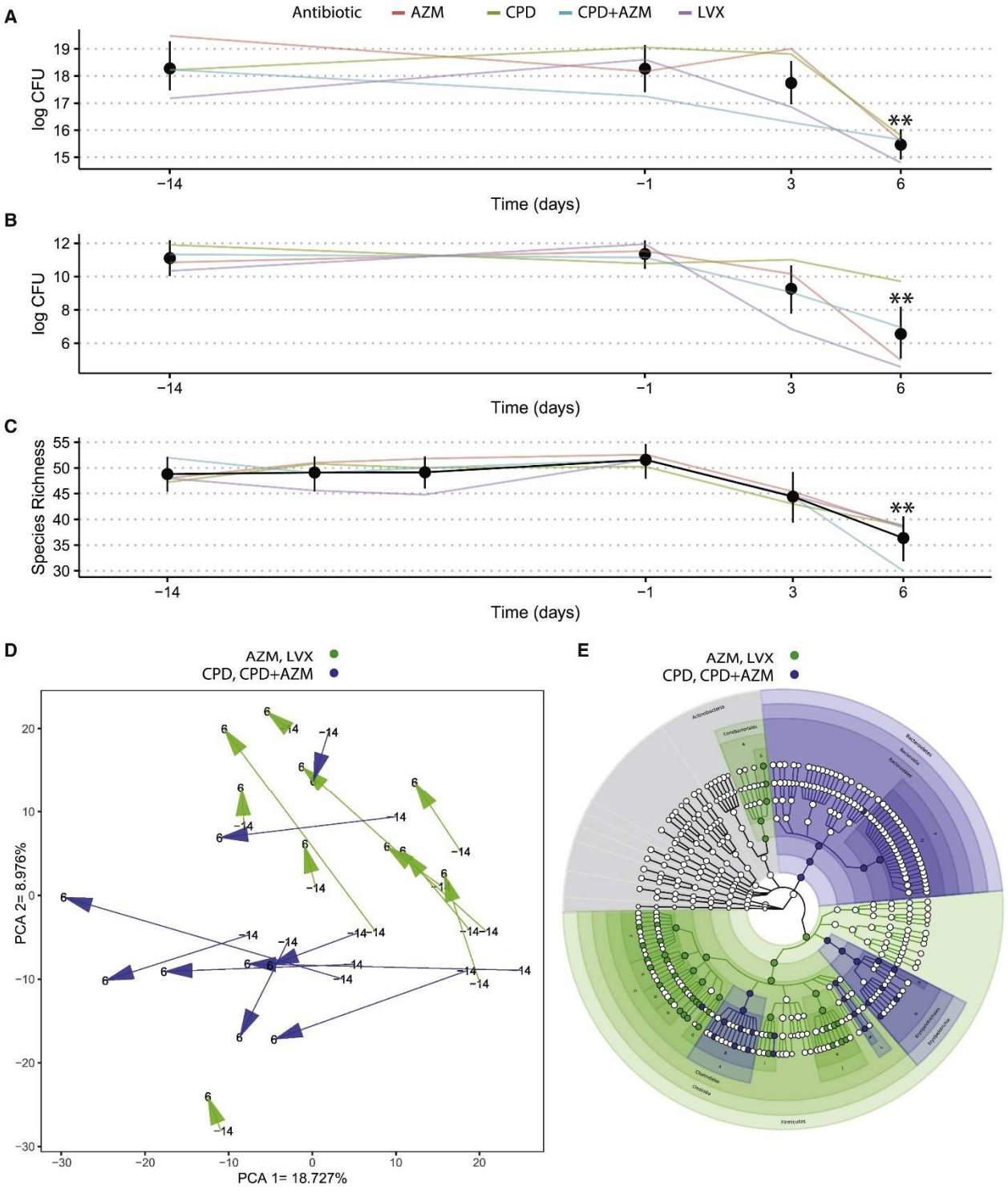


Figure 2. 1 Healthy microbiomes are perturbed by antibiotic treatment. A. Longitudinal tracking of changes in semi-quantitative culturing of aerobic bacteria. All confidence intervals are bootstrapped 95% CI of the mean. Asterisks represent significantly different timepoints (See Tables 2.3S–2.5S for p values). B. Longitudinal tracking of changes in semi-quantitative culturing of anaerobic bacteria. All confidence intervals are bootstrapped 95% CI of the mean. Asterisks represent significantly different timepoints (See

Tables 2.3S–2.5S for p values). C. Longitudinal tracking of metagenomic species richness. All confidence intervals are bootstrapped 95% CI of the mean. Asterisks represent significantly different timepoints (See Tables 2.3S–2.5S for p values.) D. The trajectory of the stool samples through the PCA space before and at the end of antibiotic administration. The PCA was generated using all of the sample points, with arrows illustrating volunteer microbiome movement through the PCA space starting at day −14 to day 6. Samples are color-coded by antibiotic—blue for CPD or CPD + AZM and green for AZM or LVX. (E) A cladogram overlaying colors representing significant effect sizes found at each taxonomic level (blue for CPD/CPD + AZM and green for AZM/LVX). See Table S7 for the biomarker legend. Each node denotes a taxonomic unit within the bacterial hierarchy, and when colored, were found to be significantly enriched in the treatment group represented by that color. White nodes represent taxonomic units that were present, but not significantly enriched in either group. See also Figures 2.8S and 2.9S.

No significant differences were found between culture results from samples taken on days -14 and -1. The day after the end of treatment, bacterial species richness also decreased significantly when compared to day -14 (11.50 CI [7.99−16.00], p = 0.0052) (Figure 2.1C, Table 2.5S). Again, there were no significant differences in species richness between any of the other three pre-antibiotic timepoints and the timepoint at day -14.

In addition to these broad trends, we also observed antibiotic-specific effects on the microbiome. Principal component analysis (PCA) using the Bray-Curtis dissimilarity of species abundance for the full dataset (all volunteer timepoints and ICU group, Figure 2.1S) was used to visualize the effect of antibiotic administration on volunteer microbiomes. We observed that volunteer microbiomes given specific treatments shared similar trajectories through PCA space after treatment, with the CPD and CPD+AZM treatment groups separating significantly from the LVX and AZM groups (p = 0.01 [PERMANOVA]) (Figure 2.1D, Figure 2.9S). Linear discriminant analysis (LDA) between the two groups at day 6 revealed an enrichment of the Bacteroidetes phylum (log average: 5.82 LDA effect size = 5.02, $p < 0.01$) and the Clostridium genus (marked as B on figure) (log average: 4.43, LDA effect size = 3.99, $p < 0.05$) in the healthy volunteers given CPD or CPD+AZM. Volunteers given LVX or AZM were enriched for genera in the Firmicutes phylum (log average: 5.73, LDA effect size = 5.03, $p < 0.01$) such as Eubacterium (log average: 5.28, LDA effect size = 4.82, $p < 0.01$), Ruminococcus (log average: 4.73, LDA

effect size = 4.34, $p < 0.01$), and Anaerostipes (log average: 3.22, LDA effect size = 3.91, $p < 0.05$) (Figure 2.1E). Notably, significant compositional differences between the groups were abrogated by day 185 ($p = 0.98$ [PERMANOVA], Figure 2.9S).

2.2.2 Azithromycin delays recovery of species richness and is associated with the relative abundance of 9 gut commensals and metabolic pathways

After treatment with antibiotics, viable aerobic and anaerobic bacteria cultured from volunteer fecal samples significantly decreased in concentration. We observed lower aerobic CFU/mL counts from fecal samples from days 6 (4.78 log-transformed CFU CI [2.89–6.36], $p = 0.0074$ [adjusted Wilcoxon rank sum]) and 12 (4.56 log-transformed CFU CI [2.47–6.86], $p = 0.0074$ [adjusted Wilcoxon rank sum]) than those recorded on day -14 (Table S3). Anaerobic CFU/mL counts exhibited a similar perturbation after treatment with reduced CFU counts on day 6 (2.90 CI [1.91–3.91], $p = 0.0047$ [Wilcoxon rank sum]) and 12 (2.83 CI [1.70–4.31], $p = 0.0047$ [Wilcoxon rank sum], Table S4). Recovery to baseline was reestablished by day 19 after treatment for both aerobic and anaerobic culture. Species number for all volunteers recovered by day 19 (table S5). Microbiome diversity recovered in a treatment specific manner, with volunteers given LVX or CPD recovering 12 days after antibiotic administration (3.00 CI [-9.99–10.00], $p = 0.4$ [Wilcoxon rank sum]) (Figure 2.2A-B).

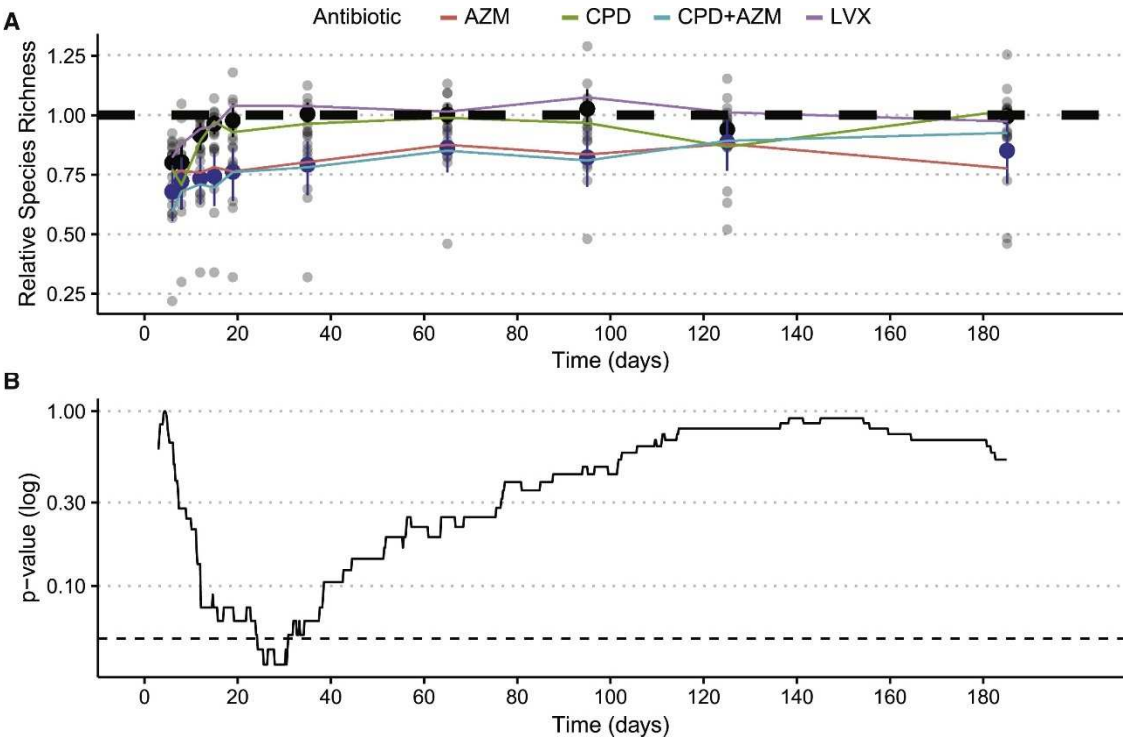


Figure 2. 2 The microbiome recovers but is delayed by AZM. A. Longitudinal recovery of relative species richness after AP. Black confidence intervals in metagenomic data represent bootstrapped 95% confidence intervals for the average of CPD and LVX volunteers at each time point, and blue confidence intervals represent 95% bootstrapped confidence intervals for AZM and XPD + AZM volunteers at each timepoint. The dashed line represents the pre-treatment average species richness. B. p value significance tests over imputed intervals between LVX and CDP and the AZM and CPD + AZM groups after antibiotic administration. The dashed line represents 0.05 p value significance threshold. See also Figures 2.10S and 2.11S.

In contrast, volunteers given AZM or CPD+AZM appeared to deviate from the recovery shown by LVX or CPD volunteers.

We discovered that AZM and CPD+AZM volunteer microbiomes, hereafter referred to as the slow recovery group (SRG), exhibited a significant delay in recovery between day 24 to 30 when compared to the microbiomes of volunteers given LVX or CPD, referred to as the fast recovery group (FRG) ($p < 0.05$) (Figure 2.2B). The average species richness of the AZM and CPD+AZM groups continued to be lower than pre-antibiotic levels until day 65 (5.00 CI [2.20e-05−1.49e01], $p < 0.051$ [Wilcoxon rank sum], Figure 2.2A-B). Eight commensal bacteria were identified to

have recovered significantly slower in volunteer microbiomes given AZM or CPD+AZM than in microbiomes given LVX and CPD: Alistepes putridensis, Bifidobacterium longum, Collinsella aerofaciens, Eubacterium eligens, Dorea longicatena, and Barnseilla intestinihominis ($p < 0.05$ [ZIBR]). The average abundance of Methanobrevibacter smithii exhibited two short blooms on days 12 and 35 in AZM and CPD+AZM. Bacteroides massiliensis was the only species identified as being enriched in AZM and CPD+AZM after treatment ($p < 0.05$ [ZIBR]).

Functional analysis confirmed that the slow recovery of Bifidobacterium longum was associated with the decreased abundance of the alternative, non-mevalonate 2-C-methyl-d-erythritol 4-phosphate/1-deoxy-d-xylulose 5-phosphate pathway (MEP/DOXP pathway), a metabolic pathway belonging to the Bifidobacterium genera.

Similar to the recovery over time in species number, the cumulative distance through the first 10 dimensions of the PCA did not differ for both the FRG and SRG. However, the SRG had a significantly higher distance through 10-dimensional Euclidean distance during recovery ($p = 0.022$, t-test on log-transformed distances, Supplemental Figure 2.10S).

2.2.3 The resistance reservoir increases in healthy volunteer microbiomes over time after antibiotic perturbation

After AP, AR gene burden increased significantly for the volunteers receiving CPD, AZM, and CPD + AZM treatments when compared to the average of their pre-antibiotic sampling points (Figure 2.3A-B, Table 2.6S).

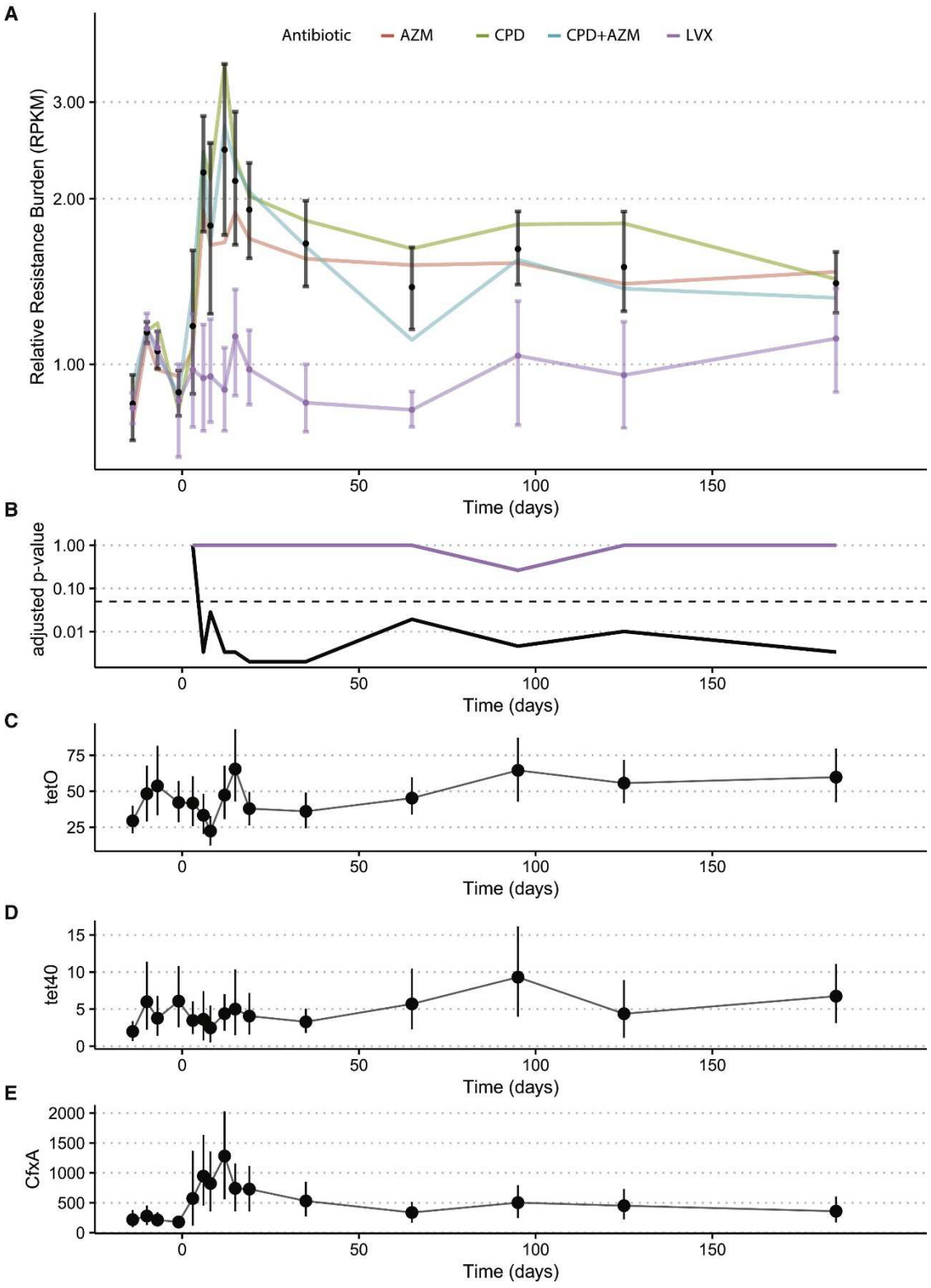


Figure 2. 3 Resistance gene burden increases up to 2 months after antibiotic perturbation. A. Resistance gene burden measured as total reads per kilobase of transcript, per million mapped reads (RPKM) of all resistance markers increases significantly immediately after antibiotic perturbation and then again at day

65. All confidence intervals are bootstrapped 95% CI of the mean. B. p value compared to first sample point at all time points. Long-term increases in tetracycline resistance after all of the treatments. The dashed line represents the 0.05 p value threshold (see Table S6 for p values). The black line represents the p value over time for AZM-, AZM+CPD-, and CPD-treated volunteers. The purple line represents the p value over time for LVX-treated volunteers. C–E. Average RPKM with bootstrapped 95% confidence intervals of the genes found to be significantly increased over time. The y axis represents average RPKM.

Conversely, volunteers receiving LVX exhibited no significant changes over time (adjusted $p < 0.05$, [paired Wilcoxon ranked sum]). Only 3 resistance elements were identified as changing significantly over time in the CPD, AZM, and CPD + AZM treatments. Three resistance genes, cfxA, tetO, and tet40, increased significantly (adjusted $p < 0.05$ [Splinectomer], Figure 2.3C-E). tetO and tet40 are often found in Firmicutes (47, 48), and we observed a concomitant increase in the average relative abundance of Firmicutes for all volunteers beginning at day 65 (Supplemental Figure 2.11S).

2.2.4 The healthy volunteer resistome after perturbation is distinct from the resistome of ICU patients

We next identified significant differences in AR gene content between the healthy volunteer microbiomes the day after the end of antibiotic treatment (day 6) and microbiomes from 26 ICU patient fecal samples (Figure 2.4A).

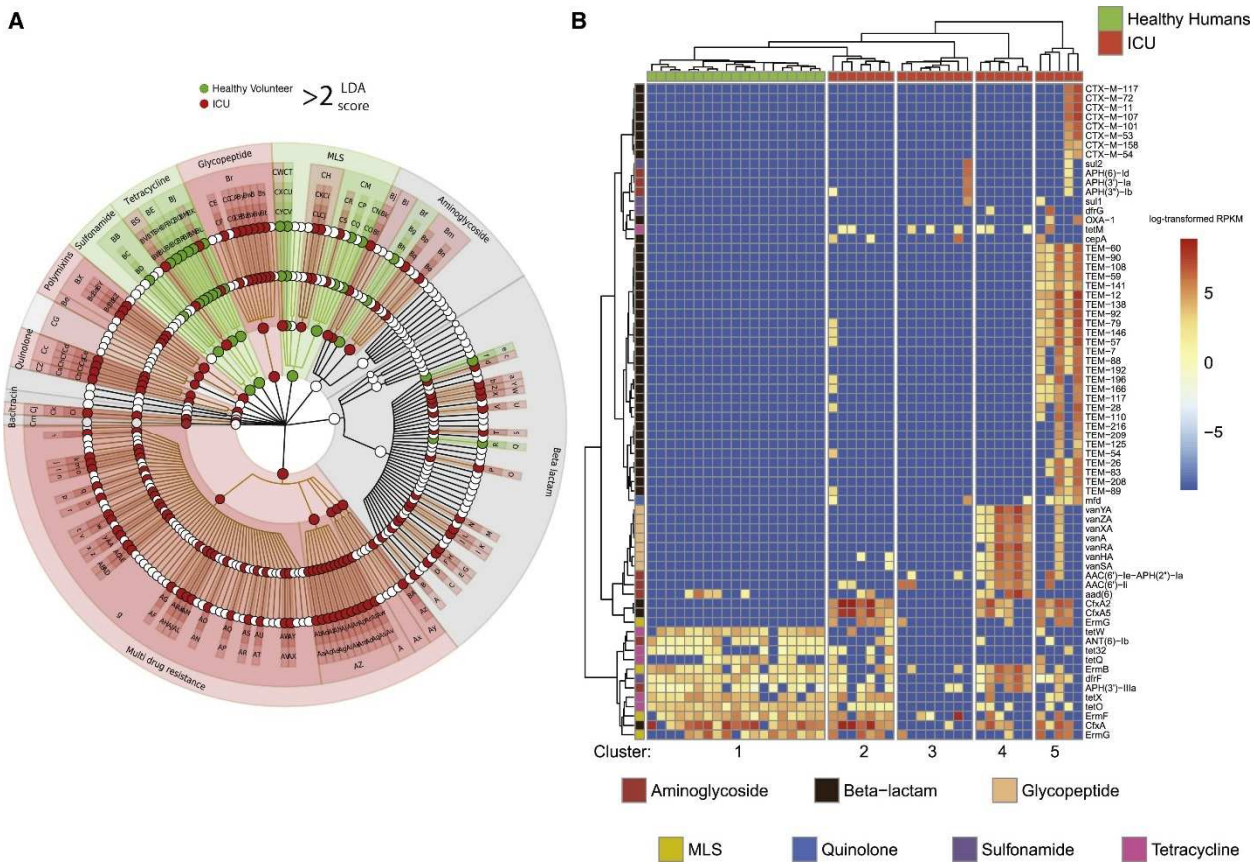


Figure 2. 4 The healthy volunteer resistome after AP is distinct from the ICU patient resistome. A. Cladogram of the antibiotic resistance genes found to be discriminatory between healthy volunteers at day 6 and ICU remnant microbiomes. The ICU microbiomes were highly enriched for multi-drug resistance and efflux pump complexes. The healthy microbiomes were enriched for cfxA resistance and the 23S ribosomal RNA methyltransferase mechanism of macrolide resistance (Au). See Table S8 for the biomarker legend. B. Heatmap of the 70 most highly variable resistance markers for the same sample set. The healthy volunteer samples cluster together and are largely depleted of the markers representative of the ICU microbiomes. The ICU microbiomes have 4 distinct clusters, which were dominated by distinct sets of variable markers. The scale represents log-transformed RPKM values.

The microbiomes of the ICU patients were characterized by enrichment of AR genes encoding multi-drug resistance (log average: 5.07, effect size = 4.75, $p < 6.46e\text{-}07$ [LDA]), specifically efflux pump classes such as ABC class efflux (log average: 3.30, effect size = 3.13, $p < 0.009$ [LDA]) and SMR class efflux (log average: 3.26, effect size = 3.14, $p < 0.009$ [LDA]). Healthy volunteer microbiomes were enriched for the beta-lactam resistance gene cfxA (log average: 5.57, effect size = 5.12, $p < 0.001$ [LDA]) and macrolide resistance genes acting on the

23S ribosomal RNA methyltransferase (log average: 5.42, effect size = 4.59, $p < 0.02$ [LDA]), both targeting antibiotic treatments just administered.

We identified the 70 most highly variable and abundant AR genes in an averaged acute post-antibiotic healthy volunteers and the ICU dataset. Hierarchical clustering identified 5 clusters, with the healthy volunteers forming one distinct cluster (Figure 2.4B). The ICU cohort was split into 4 distinct groups: cluster 3 was largely depleted of the set of highly variable genes; cluster 2 shared multiple genes overlapping with the averaged acute post-antibiotic healthy volunteer profile; cluster 4 contained glycopeptide resistance genes which confer vancomycin resistance commonly found in vancomycin-resistant Enterococcus (VRE); and finally, cluster 5 contained genes encoding mostly beta-lactam efflux pumps

2.2.5 Most healthy volunteers remain inside healthy PCA space; some enter ICU PCA space after antibiotics

To compare the effect of AP on healthy volunteer microbiomes with the presumed dysbiotic microbiome state of the ICU patients, healthy volunteer metagenomes were mapped through the first two dimensions of the PCA ordination space of Bray-Curtis dissimilarities at days -14, 6, and 185 (Figure 2.5A).

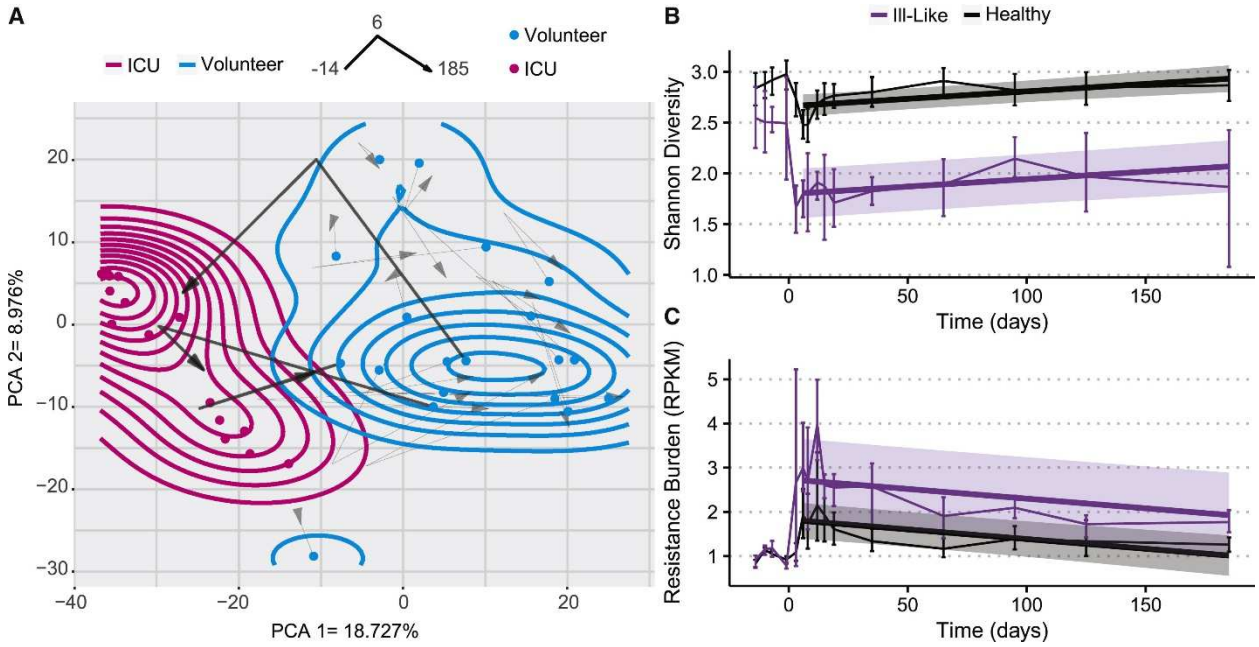


Figure 2. 5 A subset of ill-like individuals were heavily perturbed during the study. A. A 6-month longitudinal analysis of healthy volunteer microbiomes through the PCA space. The red dots are the ICU remnant microbiomes, while the blue dots represent the starting locations of all of the volunteer microbiomes. The arrows represent passage over time of the microbiomes, with the starting point samples at day −14, the inflection point is their location at day 6, and the arrowhead is their location at day 185. All of the volunteer trajectories are present, but the 3 that ended in ICU space are bold. The blue density contour was estimated using the starting coordinates of the volunteer microbiomes and overlaid onto the PCA space, and the magenta contour represents the density of the ICU microbiomes. B. Longitudinal analysis of overall resistance burden for the ICU-like subset of 3 healthy volunteers (represented in purple) and the rest of the volunteers. The error bar confidence intervals represent 95% bootstrapped confidence intervals of the mean, and a linear model estimate was fit to both groups starting after the end of antibiotic administration. Shaded regions represent 95% confidence interval for the linear fit for each group. C. The changes over time in species diversity for the ICU-like volunteers and the rest of the volunteers, with a linear model fit to both starting after the end of antibiotic administration. The error bar confidence intervals represent 95% bootstrapped confidence intervals of the mean. Shaded regions represent 95% confidence interval for the linear fit for each group.

Pre-antibiotic samples were used to estimate a density contour within the first two axes of the PCA, which we defined as the area of the pre-antibiotic "healthy state". Another density contour was estimated from ICU microbiomes to characterize the dysbiotic state of critically-ill patients. After perturbation, healthy volunteer microbiomes experienced a shift in their taxonomic composition towards a diversity minimum, where the majority of ICU patients were clustered (Figure 2.5A). Most volunteers (17 of 20) then showed a reversal of this trajectory, returning to an area near where they began; most never leaving the "healthy state" area.

However, three volunteer microbiomes experienced large decreases in diversity and richness, traversed long paths through PCA space over time, and entered the density contour demarcated by critically-ill ICU microbiomes. The PCA distance had no positive enrichment of any functional pathways. However, of the top 20 most significant negatively enriched pathways, 18 were from the genus Eubacterium. While the other volunteers lost on average 10.28 species on day 6 after antibiotic exposure, the 3 "ill-like" volunteers lost an average of 23.67. For these 3 individuals, the recovery of Shannon's diversity index, was incomplete and was significantly different from the rest of the volunteers (linear regression, $p < 4.66e-06$). By the end of the study, they had only 63.4% of the species they started with, as opposed to 99.1% recovery in the rest (Figure 2.5B). These individuals contained significantly lower abundance of 9 metabolic pathways originating from the Eubacterium genera and were instead enriched for lysine biosynthesis pathways from the genera Clostridiodes and Erysipelotrichaceae. Interestingly, their resistance gene burden did not differ from the rest of the volunteers (Figure 2.5C). This analysis indicates that a subset of the healthy volunteers was at greater risk of AP of their microbiome taxonomic diversity, and their return to a healthy state was incomplete.

2.2.6 The taxonomy and resistance composition of the healthy volunteer microbiome is altered after recovery

We next wanted to ascertain whether post-antibiotic selection resulted in effects large enough to significantly alter the composition of the microbiome, similar to what was observed with diversity. The compositional difference of the first (pre-perturbation) healthy volunteer microbiome sample was compared via Bray-Curtis dissimilarity to all the other time points from the same volunteer. Bray-Curtis dissimilarity between all pre-antibiotic timepoints was low, and no timepoints were significantly different from the starting composition (Figure 2.6A-B).

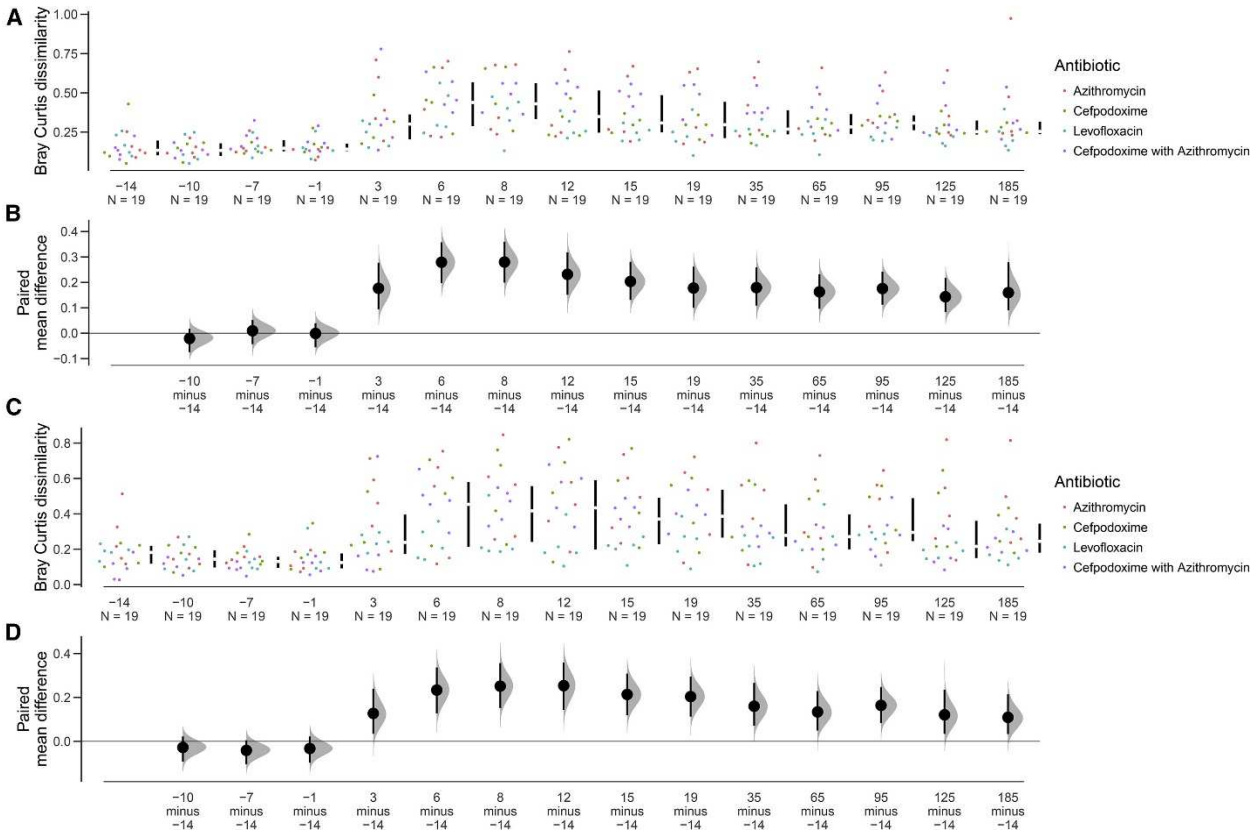


Figure 2. 6 Antibiotic perturbation in the healthy volunteer microbiome. A and C. Bray-Curtis dissimilarity of each time point to the median of the pre-antibiotic composition. Within-patient Bray-Curtis dissimilarity for all healthy volunteers was compared to the first time point. All time points for the pre-antibiotic period exhibit low dissimilarity, but this increases immediately starting with antibiotic administration and remains high until the end of the study (day 185). This analysis was conducted for species composition (A) and for resistance gene composition (C). B and D. Bootstrapped 95% confidence intervals generated for data at each time point.

However, after the start of antibiotic treatment, the Bray-Curtis dissimilarity between a volunteer's microbiome samples to their first sample increased dramatically and remained elevated for both taxonomy and AR genes (Figure 2.6C-D). This contrasts with microbial richness, which by day 40 was no longer significantly lower than pre-treatment levels (Figure 2.2B).

2.3 Materials and Methods

2.3.1 Resource Availability

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Dr. Jennie H. Kwon (j.kwon@wustl.edu).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

- Sequence data is stored as short-read metagenomic sequences in the Sequence Read Archive (SRA) and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

2.3.2 Experimental Model and Subject Details

Institutional Review of Sample Collection

Stool specimen collection was reviewed and approved by the Human Research Protection Office of Washington University in St. Louis under IRB ID #: 201610071. ICU fecal specimens was reviewed and received a non-human subjects determination by the Institutional Review Board of Washington University in St. Louis IRB ID #: 201509022.

Subject recruitment and microbiome sampling

Twenty healthy adults from the St. Louis, MO metropolitan area were recruited for the study. Subjects were eligible for the study if they were between the ages of 21 and 60 and provided

written, informed consent (Table S1). Subjects were not eligible if they met any of the following exclusion criteria: history of allergic reaction or contraindication to any of the study antibiotics; not able to provide regular fecal samples; any systemic antibiotic exposure in the previous 6 months; tube feeds in the previous 6 months; pregnant, risk of becoming pregnant, or breastfeeding during the study period; gastroenteritis in previous 3 months; non-elective hospitalization in previous 12 months; incontinent of stool; known colonization with a MDRO; anticipated change in diet or medications during the study period; elective surgery planned during the study period; history of an intestinal disorder; or inability to provide written, informed consent. Care was taken to recruit a volunteer cohort with equal sex parity (see Table S1). After enrollment, subjects were interviewed about their medical history and underwent a physical exam. Subjects were randomly assigned to receive a 5-day course of one of the following antibiotics: AZM, LVX, CPD, or CPD+AZM. They were directed to begin the antibiotic on the appropriate day. Subjects sent back their empty pill bottles to confirm that all doses were taken. Study personnel involved in patient recruitment, specimen collection and processing were blinded to treatment group.

Study personnel provided subjects with supplies for collecting clinical specimens. Subjects submitted fecal specimens to study personnel at 15 time points throughout the study period (Table S2). Fecal samples from pre-specified time points underwent microbiologic culture upon receipt of the specimen (Table S2). All other specimens and remnant fecal specimens were frozen on the same day they were provided. At each specimen submission point, subjects also completed questionnaires on bowel movement consistency and frequency, diet, medications, and changes in medical history. This study received approval from the Washington University Human Research Protection Office. The ICU fecal specimens were convenience samples of

remnant stool from the Clinical Microbiology Laboratory collected from 26 ICU inpatients who had diagnostic testing for *Clostridioides difficile* at Barnes-Jewish Hospital, thus no clinical metadata was collected.

2.3.3 Semi-quantitative culturing

To provide an overall assessment of the aerobic and anaerobic microbial burden in the samples serial dilutions were plated to general purpose aerobic and anaerobic culture media. For aerobic culture, ~1g or 1ml of fecal waste was placed in a 2ml Nunc tube. An equal amount of 1X PBS (1:1 dilution) was added and vortexed well to homogenize the sample. 10, 10-fold dilutions into PBS with sample were then made. BAP plates (Hardy Diagnostics) were inoculated with 10µl and 100µl of each dilution respectively; a cross streak pattern was used for the 10µl plate and a quadrant streak was used for the 100µl plate, and the plates were incubated in air overnight at 35°C. After incubation, colonies on the 10µl plate were counted, and anything over 50 was recorded as ">50 colonies". For the 100µl plate a semi-quantitative amount of growth was recorded (0,+1,+2,+3,+4) and a general physical description of the colonies was recorded (e.g. mixed enteric flora, mixed gram positive flora, etc.). Sweeps of the plate were frozen in molecular grade water and in TSB-glycerol. Samples were then stored at -80°C.

For anaerobic culture, a dilution series was created from fecal samples as just described. Similar aliquots of dilutions were inoculated onto 2 pre-reduced BBA (Hardy Diagnostics) plates and incubated at 35°C in an anaerobic chamber or anaerobic bag overnight. Streaking methods were the same as just described.

2.3.4 DNA extraction and sequencing

Fecal samples were kept on dry ice while samples were chipped off and weighed to standardize the amount of fecal matter to be extracted. Total gDNA was extracted from samples using a

PowerSoil DNA isolation kit (MoBio). Lysis was conducted using a Minibeadbeater (Biospec Products). gDNA was stored at -20°C until sequencing library creation. Shotgun metagenomic sequencing libraries were created using gDNA diluted to 0.5 ng/µl and the modifications to the Nextera library prep lot (Illumina) detailed in Baym et al. (49). This generated ~450 bp DNA fragments which were purified using the Agencourt AMPure XP system (Beckman Coulter) and quantified using the Quant-it PicoGreen dsDNA assay (Invitrogen). Samples were pooled onto lanes to ensure ~3M (2x150bp) reads per sample, and three technical replicates of this pooling process were quantified using the Qubit dsDNA BR assay and then combined in equimolar concentrations to reduce stochastic error in read distribution during the sequencing process. Samples were sequenced in Illumina NextSeq High-output sequencing machines at the Edison Family Center for Genome Sciences and System Biology at Washington University School of Medicine in St. Louis.

2.3.5 Processing of Illumina sequence data

Total reads were demultiplexed by barcode into individual sample sequence bins. The adapters and barcodes were removed by Trimmomatic and simultaneously filtered for quality. Human reads were filtered using DeconSeq by mapping reads to the human genome (GrCh38) (50, 51). A minimum sample read depth of 2.5M was determined using species and community resistance rarefaction curves generated with MetaPhlAn2 and ShortBRED (52, 53).

2.3.6 Quantification and statistical analysis

Community taxonomy and resistome quantification and downstream analysis

Species composition was predicted using MetaPhlAn2 (54). Resistome composition in RPKM (reads per kilobase of reference sequence per million sample reads) was estimated using ShortBRED based off a custom-built marker database including the CARD database (52).

Functional output of the microbiome was estimated using HUMAnN 2 (55) and significant pathway enrichment was conducted using MaAsLin2 (56). Species richness and diversity were calculated using the R package vegan (version 2.5-6). Abundance and prevalence thresholding was conducted to exclude low abundance, and thus low confidence, species. We used an abundance threshold of .01 % and a prevalence threshold of 20% (of all healthy volunteer samples) and 19% (of all ICU samples) to result in a median loss of less than 10% overall relative species abundance abundance in all taxonomic samples. Lefse analysis was conducted on the full, unthresholded metaphlan taxonomic abundance tables.

For the resistance analysis, we used an abundance threshold of .5 % and a prevalence threshold of 20% (of all healthy volunteer samples) and 5% (of all ICU samples) resulting in 70 highly variable genes across both datasets. Significance for longitudinal data analysis of semi-quantitative culture, taxonomy, and resistome analysis was conducted using the paired Wilcoxon rank sum test in R. Loess fitting of species richness recovery for antibiotic treatments was estimated using the base r loess function, and 100 intervals were then imputed along the fit. The area under the curve was compared to a null distribution made from 999 random permutations of the data using the SplinctomeR package (57). Increases in specific resistance genes was tested similarly using the SplinectomeR package, comparing the loess regression of a specific gene to a null distribution computed from regressions of all other genes over time. Significance tests begin after first 5% of the regression to account for flying tails. Patient samples were normalized to an average of their pre-antibiotic samples for each metric studied. 2D kernel density estimation of ill and healthy density contours in PCA space were estimated using the MASS (version 7.3.51.6) R package and visualized in ggplot2 (version 3.3.1). Principle component analysis was conducted without Patient 2, as they did not submit most of their samples. LDA effect size

analysis was conducted in LEfSe (58) using standard parameters, antibiotic as a class, per sample normalization, and one-against-all multiclass analysis. Results visualized using using LEfSe (Galaxy module) and GraPhlAn v0.9, respectively (58, 59). Commensals significantly associated with AZM delayed microbiome restoration were identified using a zero-inflated two part mixed effects model using the zibr (version 0.1) R package (60). Zibr requires a complete dataset (no missing datapoints) so missing datapoints (5 of 209 datapoints) were imputed using the average abundance of each species for the entire dataset. Patient 2, which did not submit most of the fecal samples, was not included. Heatmap of differences in ICU vs. healthy volunteer resistome was created using the pheatmap (version 1.0.12) package in R using the 71 most highly variable, log-transformed AR gene abundances in an averaged acute post-antibiotic healthy volunteer profile (three time points immediately after cessation of antibiotics) and the ICU dataset, after filtering to remove efflux pumps, low abundance markers, or genes not targeting one of the major antibiotic categories. Patient 2 was also removed from this analysis due to lack of samples. Cluster analysis of samples used the ward.D2 clustering algorithm (61) on the Euclidean distance computed from log transformed RPKM abundances of each resistance gene for all samples and then sorted using the dendsort (version 0.3.3) R package. The longitudinal species diversity of the microbiomes of the three patients who transitioned to an ill-like microbiome state was fit to a linear model and compared to the rest of the healthy patients; visualization was created using ggplot2 (version 3.3.1). Longitudinal analysis of taxonomic and community resistance composition of within-patient bray-curtis (computed using R package vegan version 2.5-6) sample dissimilarity was conducted by first normalized to an average of their pre-antibiotic microbial compositions. Then the paired mean difference of bootstrapped confidence intervals between all timepoints was compared to the first using the dabestr (version 0.2.5) R package

(62). Sequence data is stored as short-read metagenomic sequences in the Sequence Read Archive (SRA); Healthy volunteer data is stored in bioproject PRJNA664754 while the ICU metagenomes are stored in bioproject PRJNA703034.

2.4 Discussion

Antibiotic exposure may result in acute and persistent changes in the commensal host microbiome (3-5, 10) . We found significant reduction of viable bacterial titers in aerobic and anaerobic microbiologic culture, and a reduction in metagenomic species richness. These results are consistent with previous studies involving antibiotics in healthy individuals and support our initial hypothesis that antibiotics can cause acute perturbations in the gut microbiome of healthy volunteers (6, 63). This reduction was first observed on day 6, the first day after the 5-day treatment. This was likely due to our study design, which collected timepoints the first day of treatment, before the antibiotic perturbation had time to reach the stool. The small number of volunteers and 4 different antibiotic treatment groups may have included too much variation to detect the beginning of decreased species richness and colony forming units at the next sampling point, which was day 3 of the 5-day treatment. Palleja et al. found a significant decrease in metagenomic operational taxonomic units 1 day after the end of a four-day treatment with a 3-drug cocktail, which correlates well with our observation of a significant decrease one day after the end of the 5 day treatment in our study (6). Given the information from this previous study, it seems reasonable that had we collected samples on day 4 of the 5-day treatment, we may have seen a significant difference in species richness or culture CFUs.

Furthermore, we observed a direct relationship between species relative abundance and resistance elements. Immediately after treatments containing CPD, there was a relative

enrichment of Bacteroidetes. Bacteriodes spp. have well documented resistance to beta-lactams, and the cfxA gene produced phenotypic resistance to cefoxitin in Bacteriodes species (64-66). cfxA increased in relative abundance only after CPD or CPD+AZM treatment, suggesting this is a primary form of resistance for enteric Bacteriodes spp. as well. Previous studies had found that Cefprozil, a second-generation cephalosporin, induced increases in Enterobacter cloacae in healthy volunteers with low diversity, Bacteroides dominated microbiomes (67). Our results indicate that Bacteroides species survive CPD treatment, likely via cfxA, resulting in a low diversity, high Bacteroides environment, generating opportunity for expansion of pathogens such as Enterobacter spp. In comparison, the AZM and LVX groups were enriched for multiple gram-positive genera within the Firmicutes phylum. The increases correlates well with observed increases in the Bacteriodes/Firmicutes ratio in mice (68). These taxonomic changes were implicated in increased adipogenesis, altered microbiome short-chain fatty acid production, and other risk factors for obesity.

Recovery of species richness after exposure varied by treatment, similar to previous studies (6, 69). The SRG, which consisted of volunteers given AZM or AZM+CPD experienced a 6-day period of extended lower diversity, approximately the duration of the terminal half-life for this drug of 5 days (70). This is in sharp contrast to the much shorter half-life of CPD and LVX, which are reported to be 2-3 and 6-8 hours respectively; this could partly explain the slower recovery of species richness in the SRG (71, 72). Previous analyses using 16S rRNA gene amplicon data reported a reduction in species richness 1 year after children were treated with the macrolides clarithromycin and erythromycin, as well as a reduction in the abundance of Bifidobacterium and Collinsella genera (69). Our time series data in adults identified both B. longum and C. aerofaciens as well as E. eligens and D. longicatena recovered slower after AZM.

B. longum has been shown to be reduced in abundance in both the elderly (18, 73) and the critically ill (14, 74). A recent study found a negative correlation between increased abundance of the Bacteriodes and Methanobrevibacter genera and healthy microbiome (75), and the change over time of the two species significantly associated with slow recovery hail from these genera. The two recovery groups had similar cumulative distances through PCA space, but the SRG also had a significantly higher distance between their first and last samples. The stability of the post-antibiotic microbiome is well documented (76, 77), but the effect of time-induced drift associated with one antibiotic versus another is not well understood. Our results confirm that antibiotic treatment produces ecological opportunity in healthy individuals, reshaping even robust fecal microbiomes through increased resistance gene content and altered taxonomic composition. The increased net compositional distance exhibited by the SRG is likely due to the longer bioavailability of AZM driving prolonged niche discordance and ecological opportunity. Increased intra-patient temporal variation in stool and skin microbiomes has been significantly associated with adverse infectious outcomes after patients with acute myeloid leukemia undergoing induction chemotherapy (78). Cooption of the community instability inherent in ecological opportunity by pathogenic organisms has been theorized to drive pathogen virulence and adaptation (26). The SRG was also associated with lower counts of the Bacteriodes pentose phosphate pathway, as well as the alternative, non-mevalonate 2-C-methyl-d-erythritol 4-phosphate/1-deoxy-d-xylulose 5-phosphate pathway (MEP/DOXP pathway) from the genus Bifidobacterium (of which B. longum, one of the species identified as recovering slower in this group, is a part) (79). This pathway is also present in many pathogens (80), and is currently being investigated as a target for antibiotic development due to its absence in humans and the known ability of fosmidomycin to inhibit DCP reductoisomerase, a key enzyme in the pathway (81). It

is hypothesized to be a method to modulate host response and lower the chance of an immunogenic reaction to commensal bacteria (82).

When we compared antibiotic resistance composition, we found that immediately after a 5-day course of antibiotics AR gene burden in healthy microbiomes increased and remained elevated for the length of the study in volunteers receiving CPD, AZM, and CPD + AZM. This initial increase in AR gene burden was not unexpected, and similar increases have been described (6). The length of elevated AR burden varied from previous studies; Palleja et al. reported no significant increases in total AR burden by 45 days in adults (6), however D'Souza et al. reported significant increases in AR prevalence by 6 months in HIV-exposed, uninfected infants receiving weekly clotrimoxazole prophylaxis (12). The reason for these differences in severity and duration of AR elevation is likely to be governed by the maturity, stability, and health of the microbiomes under study, and the effects of each specific antibiotic treatment, as LVX did not increase resistance gene burden.

While AR gene burden was in general higher in ICU patient microbiomes, certain classes of AR genes in healthy volunteers were enriched compared to ICU patient microbiomes including AR genes classes specific to the antibiotics given during the study, such as the beta-lactam resistance gene cfxA (CPD) and 23S ribosomal RNA methyltransferase resistance (AZM). ICU patients were instead enriched for multi-drug resistance. In the generalist-specialist game theory model of ecological succession, generalists initially dominate environments with high ecological opportunity during population expansion (29). Remodeling of the healthy individual resistome after antibiotic exposure resulted in increases in three genes, tetO, cfxA, and tet40, two of which do not convey resistance to any of the antibiotics given in the study, confirming that antibiotic perturbation creates opportunities for species with generally broad antibiotic resistance to

dominate transiently. While some AR genes can reliably be used as signatures of specific bacteria, such as ampC for Enterobacteriaceae (83, 84), tetO may instead be an indicator of previous antibiotic exposure and subsequent microbiome-wide increase in AR gene burden and selection (85). Short courses of antibiotics could trigger the acquisition or entrenchment of diverse resistance genes, leading to increased AR seen in some ICU patients and the elderly (74, 86). The persistent increases in overall AR burden, in tandem with reduced microbiome diversity and similarity to ICU microbiomes of a subset of volunteers, could be used as biomarkers for future adverse reactions to antibiotic treatment, or for higher risk of hospital acquired infection as has been proposed for other diseases (87).

Though antibiotics caused an average decrease in species richness, volunteer microbiomes returned to a pre-antibiotic level within 2 months. This resilience is similar to previous studies which reported antibiotic exposure yielded only a short-term effect on species richness (4, 20). Despite this general trend of resilience, the microbiome of 3 individuals recovered slower than the rest of the volunteers. Instead they exhibited substantial movement over time through PCA space, ending the observation window within the PCA space dominated by ICU patient microbiomes, and had significant changes to their metabolic output, with reduced pathway abundance from the Eubacterium genera, an important gut commensal, but increased abundance of pathways from the genus Erysipelotrichaceae. Erysipelotrichaceae species are known to be highly immunogenic (88), linked to inflammatory bowel disease (IBD) and increase in relative abundance post-antibiotics (89). The PCA distance between the first and last sample of a volunteer was negatively correlated with 18 Eubacterium metabolic pathways, which suggests that functional pathways originating from that genus could be important to reducing the large changes over PCA space that the 3 ill-like volunteers underwent. The concept of “antibiotic

scarring" has been used in previous work (8) to describe long term AR gene accumulation in pediatric microbiomes; we propose a modified definition wherein scarring is characterized as a significantly altered, perturbed taxonomic composition with increased AR burden (a generalized response containing on- and off- target resistance elements) after antibiotic exposure. We further identify that antibiotic scarring pushes some low diversity microbiomes towards an ill-like phenotype, demonstrating that long-term microbial community perturbation can occur from a single dose of antibiotics in healthy individuals. This definition has potential to be integrated into patient-care models optimized for selecting antibiotic treatment personalized to an individual's unique microbiome composition, an important goal of antibiotic stewardship.

2.4.1 Limitations of the study

Though this is a pilot-study and is limited in scope, we believe that it creates a usable framework for studying the universal effects of antibiotics on the healthy human microbiome, as well as for identifying microbiome compositions which may be more prone to potential antibiotic-induced dysbiosis. One limitation of short-read shotgun metagenomic DNA sequencing lies in its inability to determine what microbiota functions are actively being expressed. While our data suggest that relatively few AR genes increase in abundance long-term after AP, future studies leveraging metagenomic RNA (cDNA) sequencing or long-read sequencing could determine if this correlates with increased gene expression or horizontal transmission within the microbial community, respectively (90, 91). This study's small treatment group size and 6-month study window also limit the ability to discern whether the observed alterations to the taxonomy and resistome persist to longer intervals, though other studies have recorded similar trends years after antibiotic exposure (8, 20). Regardless, our results illustrate a dynamic but somewhat incomplete recovery process that is dependent on the specific antibiotic regiment and highlight the potential

intrinsic resilience of particular gut microbial community architectures. Strain-level dynamics during antibiotic perturbation provide an extra layer of complexity and may help explain the remarkable change in the composition of the resistome (92). The ICU comparator group were collected during routine C. difficile surveillance testing, and thus no clinical metadata were collected that could help identify patient covariates or further explain the variation in antibiotic perturbation in healthy volunteers. Though any observation of differential resistance elements between ICU patients and healthy volunteers is impossible to correlate to previous ICU antibiotic exposure without extensive knowledge of their past clinical history, this comparator group serves as an example of the low diversity, high MDR taxonomic state found in ill patients, and thus of a general dysbiosis phenotype (13).

In conclusion, our findings indicate that short courses of antibiotics commonly used for treatment of bacterial infections can cause both short- and longer-term perturbations and scarring of the microbiome in healthy human volunteers, resulting in the prolonged increase of AR in healthy microbiomes. We further refine the definition of antibiotic scarring, identifying resistance genes which could be used to detect previous perturbation through the expansion of broadly AR resistant organisms. Finally, we observed that AZM administration results in delayed recovery and greater compositional distance by the end of the study (figure 2.7).

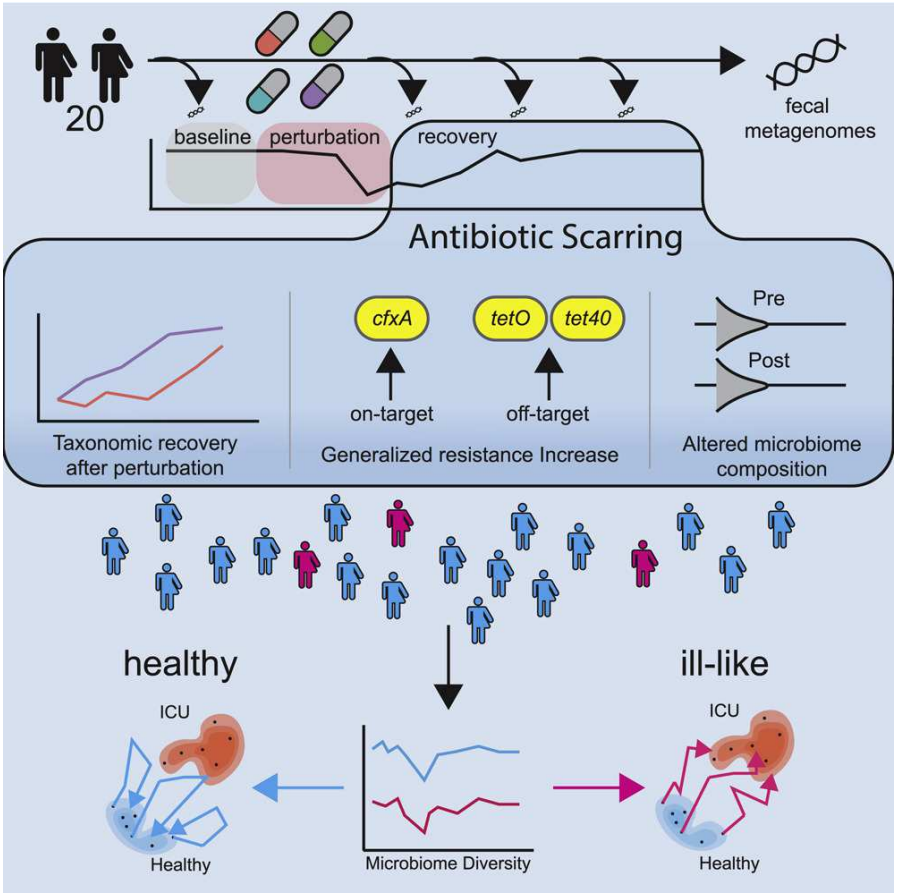


Figure 2. 7 Overview of the results of the study. Antibiotics are deployed against bacterial pathogens, but their targeting of conserved microbial processes means they also collaterally perturb the commensal microbiome. To understand acute and persistent effects of antibiotics on the gut microbiota of healthy adult volunteers, we quantified microbiome dynamics before, during, and 6 months after exposure to 4 commonly used antibiotic regimens. We observed an acute decrease in species richness and culturable bacteria after antibiotics, with most healthy adult microbiomes returning to pre-treatment species richness after 2 months, but with an altered taxonomy, resistome, and metabolic output, as well as an increased antibiotic resistance burden. Azithromycin delays the recovery of species richness, resulting in greater compositional distance. A subset of volunteers experience a persistent reduction in microbiome diversity after antibiotics and share compositional similarities with patients hospitalized in intensive care units. These results improve our quantitative understanding of the impact of antibiotics on commensal microbiome dynamics, resilience, and recovery.

Further study and interventions are necessary to mitigate the development of AR and better identify individuals at risk of developing long term negative effects after treatment. The approaches described in this work may be applicable for measuring the impact of existing and newly developed antibiotics on the gut microbiome and resistome, with such perturbation measures incorporated into antibiotic lead compound selection and development (93, 94). The

long-term increase in resistance burden observed in this study is an example of events which can push a low AR resistome towards higher AR. With the continued development of AR, novel methods to understand and prevent AR are necessary, and these data form a resource for studying both the short- and long-term effects of antibiotic perturbation on the healthy microbiome and resistome.

2.5 Acknowledgements

This work was co-authored by Winston E. Anthony, Bin Wang, Kimberley V. Sukhum, Alaric W. D'Souza, Tiffany Hink, Candice Cass, Sondra Seiler, Kimberly A. Reske, Christopher Coon, Erik R. Dubberke, Carey-Ann D. Burnham, Gautam Dantas, Jennie H. Kwon in *Cell Reports*(95). We thank the participants enrolled in this study for their time and participation. This work was supported in part by a United States Agency for International Development award (award number 3220-29047) to CAB and GD. WEA was supported by a T32 NIH Ruth L. Kirschstein National Research Training Grant Fellowship: 5T32GM007067-44. KVS was supported by the Society for Healthcare Epidemiology of America Research Scholar Award. AWD was supported by the Institutional Program Unifying Population and Laboratory-Based Sciences Burroughs Welcome Fund grant to Washington University. This work is supported by the Centers for Disease Control and Prevention (OADS BAA 2016-N-17812) Contract #200-2016-90962 to JHK. JHK is supported by 1K23AI137321-01A1 from the National Institute of Allergy and Infectious Diseases. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. The authors thank members of the Dantas lab for helpful comments on the manuscript. The authors thank the Edison Family Center for Genome Sciences & Systems Biology at Washington University School of Medicine

in St Louis staff (Eric Martin, Brian Koebbe, Jessica Hoisington-López, and MariaLynn Crosby) for technical support in high-throughput sequencing and computing.

2.5.1 Author Contributions

Conceptualization: C.D.B., E.R.D., J.H.K., and G.D.; Formal Analysis: W.E.A, K.V.S., and A.W.D.; Investigation: B.W., T.H., Ca. C., S. S., K.R., and Ch. C.; Writing – Original Draft: W.E.A.; Writing – Review & Editing: W.E.A. and K.V.S.; Supervision: J.H.K and G.D.; Resources: C.A.B, J.H.K, G.D.; Funding Acquisition: C.D.B., E.R.D., J.H.K., and G.D.

2.5.2 Declaration of Interests

The authors declare no competing interests.

2.5.3 Inclusion and Diversity

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure ethnic or other types of diversity in the recruitment of human subjects. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper received support from a program designed to increase minority representation in science.

2.6 References

1. Nicolaou KC, Rigol S. A brief history of antibiotics and select advances in their synthesis. The Journal of Antibiotics. 2018;71(2):153-84.

2. Ribeiro da Cunha B, Fonseca LP, Calado CRC. Antibiotic Discovery: Where Have We Come from, Where Do We Go? Antibiotics (Basel). 2019;8(2):45.

3. Wipperman MF, Fitzgerald DW, Juste MAJ, Taur Y, Namasivayam S, Sher A, et al. Antibiotic treatment for Tuberculosis induces a profound dysbiosis of the microbiome that persists long after therapy is completed. Scientific Reports. 2017;7(1):10767.

4. Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS biology. 2008;6(11):e280-e.

5. Ubeda C, Taur Y, Jenq RR, Equinda MJ, Son T, Samstein M, et al. Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. J Clin Invest. 2010;120(12):4332-41.

6. Palleja A, Mikkelsen KH, Forslund SK, Kashani A, Allin KH, Nielsen T, et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. Nature Microbiology. 2018;3(11):1255-65.

7. de Gunzburg J, Ghozlane A, Ducher A, Le Chatelier E, Duval X, Ruppé E, et al. Protection of the Human Gut Microbiome From Antibiotics. J Infect Dis. 2018;217(4):628-36.

8. Gasparrini AJ, Wang B, Sun X, Kennedy EA, Hernandez-Leyva A, Ndao IM, et al. Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome. Nature Microbiology. 2019;4(12):2285-97.

9. Gasparrini AJ, Crofts TS, Gibson MK, Tarr PI, Warner BB, Dantas G. Antibiotic perturbation of the preterm infant gut microbiome and resistome. Gut Microbes. 2016;7(5):443-9.

10. Stevens V, Dumyati G, Fine LS, Fisher SG, van Wijngaarden E. Cumulative Antibiotic Exposures Over Time and the Risk of Clostridium difficile Infection. Clinical Infectious Diseases. 2011;53(1):42-8.

11. Brown KA, Khanafer N, Daneman N, Fisman DN. Meta-analysis of antibiotics and the risk of community-associated Clostridium difficile infection. Antimicrob Agents Chemother. 2013;57(5):2326-32.

12. D'Souza AW, Moodley-Govender E, Berla B, Kelkar T, Wang B, Sun X, et al. Cotrimoxazole Prophylaxis Increases Resistance Gene Prevalence and α-Diversity but Decreases β-Diversity in the Gut Microbiome of Human Immunodeficiency Virus–Exposed, Uninfected Infants. Clinical Infectious Diseases. 2019.

13. McDonald D, Ackermann G, Khailova L, Baird C, Heyland D, Kozar R, et al. Extreme Dysbiosis of the Microbiome in Critical Illness. mSphere. 2016;1(4):e00199-16.

14. Ojima M, Motooka D, Shimizu K, Gotoh K, Shintani A, Yoshiya K, et al. Metagenomic Analysis Reveals Dynamic Changes of Whole Gut Microbiota in the Acute Phase of Intensive Care Unit Patients. Dig Dis Sci. 2016;61(6):1628-34.

15. Gershuni VM, Friedman ES. The Microbiome-Host Interaction as a Potential Driver of Anastomotic Leak. Current Gastroenterology Reports. 2019;21(1):4.

16. Meng M, Klingensmith NJ, Coopersmith CM. New insights into the gut as the driver of critical illness and organ failure. Curr Opin Crit Care. 2017;23(2):143-8.

17. Araos R, Battaglia T, Ugalde JA, Rojas-Herrera M, Blaser MJ, D'Agata EMC. Fecal Microbiome Characteristics and the Resistome Associated With Acquisition of Multidrug-Resistant Organisms Among Elderly Subjects. Frontiers in Microbiology. 2019;10(2260).

18. Rampelli S, Soverini M, D'Amico F, Barone M, Tavella T, Monti D, et al. Shotgun Metagenomics of Gut Microbiota in Humans with up to Extreme Longevity and the Increasing Role of Xenobiotic Degradation. mSystems. 2020;5(2):e00124-20.

19. Li J, Si H, Du H, Guo H, Dai H, Xu S, et al. Comparison of gut microbiota structure and Actinobacteria abundances in healthy young adults and elderly subjects: a pilot study. BMC Microbiol. 2021;21(1):13-.

20. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A. 2011;108 Suppl 1(Suppl 1):4554-61.

21. De La Cochetière MF, Durand T, Lepage P, Bourreille A, Galmiche JP, Doré J. Resilience of the dominant human fecal microbiota upon short-course antibiotic challenge. Journal of clinical microbiology. 2005;43(11):5588-92.

22. Mangin I, Lévêque C, Magne F, Suau A, Pochart P. Long-term changes in human colonic Bifidobacterium populations induced by a 5-day oral amoxicillin-clavulanic acid treatment. PLoS One. 2012;7(11):e50257-e.

23. Zaura E, Brandt BW, Teixeira de Mattos MJ, Buijs MJ, Caspers MPM, Rashid M-U, et al. Same Exposure but Two Radically Different Responses to Antibiotics: Resilience of the Salivary Microbiome versus Long-Term Microbial Shifts in Feces. mBio. 2015;6(6):e01693.

24. Wellborn GA, Langerhans RB. Ecological opportunity and the adaptive diversification of lineages. Ecol Evol. 2015;5(1):176-95.

25. Simpson GG. Tempo and mode in evolution: Columbia University Press; 1984.

26. Scanlan PD. Microbial evolution and ecological opportunity in the gut environment. Proc Biol Sci. 2019;286(1915):20191964-.

27. McDonald JE, Marchesi JR, Koskella B. Application of ecological and evolutionary theory to microbiome community dynamics across systems. Proceedings of the Royal Society B: Biological Sciences. 2020;287(1941):20202886.

28. Ferretti P, Pasolli E, Tett A, Asnicar F, Gorfer V, Fedi S, et al. Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. Cell Host Microbe. 2018;24(1):133-45.e5.

29. Angell IL, Rudi K. A game theory model for gut bacterial nutrient utilization strategies during human infancy. Proceedings of the Royal Society B: Biological Sciences. 2020;287(1931):20200824.

30. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. Science Translational Medicine. 2009;1(6):6ra14.

31. Cabral DJ, Wurster JI, Korry BJ, Penumutchu S, Belenky P. Consumption of a Western-Style Diet Modulates the Response of the Murine Gut Microbiome to Ciprofloxacin. mSystems. 2020;5(4):e00317-20.

32. Sekirov I, Tam NM, Jogova M, Robertson ML, Li Y, Lupp C, et al. Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. Infect Immun. 2008;76(10):4726-36.

33. Theriot CM, Koenigsknecht MJ, Carlson PE, Jr., Hatton GE, Nelson AM, Li B, et al. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. Nat Commun. 2014;5:3114-.

34. Laubitz D, Typpo K, Midura-Kiela M, Brown C, Barberán A, Ghishan FK, et al. Dynamics of Gut Microbiota Recovery after Antibiotic Exposure in Young and Old Mice (A Pilot Study). Microorganisms. 2021;9(3):647.

35. Daillère R, Vétizou M, Waldschmitt N, Yamazaki T, Isnard C, Poirier-Colame V, et al. Enterococcus hirae and Barnesiella intestinihominis Facilitate Cyclophosphamide-Induced Therapeutic Immunomodulatory Effects. Immunity. 2016;45(4):931-43.

36. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science. 2018;359(6371):97-103.

37. Park JC, Im S-H. Of men in mice: the development and application of a humanized gnotobiotic mouse model for microbiome therapeutics. Experimental & Molecular Medicine. 2020;52(9):1383-96.

38. Nguyen TLA, Vieira-Silva S, Liston A, Raes J. How informative is the mouse for human gut microbiota research? Dis Model Mech. 2015;8(1):1-16.

39. Sievert DM, Ricks P, Edwards JR, Schneider A, Patel J, Srinivasan A, et al. Antimicrobial-Resistant Pathogens Associated with Healthcare-Associated Infections Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009–2010. Infection Control & Hospital Epidemiology. 2013;34(1):1-14.

40. Shetty SA, Hugenholtz F, Lahti L, Smidt H, de Vos WM. Intestinal microbiome landscaping: insight in community assemblage and implications for microbial modulation strategies. FEMS Microbiol Rev. 2017;41(2):182-99.

41. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, et al. The long-term stability of the human gut microbiota. Science. 2013;341(6141):1237439-.

42. Huse SM, Ye Y, Zhou Y, Fodor AA. A Core Human Microbiome as Viewed through 16S rRNA Sequence Clusters. PLoS One. 2012;7(6):e34242.

43. Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, et al. Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun. 2020;11(1):5206.

44. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, et al. Infectious Diseases Society of America/American Thoracic Society Consensus Guidelines on the Management of Community-Acquired Pneumonia in Adults. Clinical Infectious Diseases. 2007;44(Supplement_2):S27-S72.

45. Kollef MH, Fraser VJ. Antibiotic Resistance in the Intensive Care Unit. Annals of Internal Medicine. 2001;134(4):298-314.

46. MacVane SH. Antimicrobial Resistance in the Intensive Care Unit: A Focus on Gram-Negative Bacterial Infections. Journal of Intensive Care Medicine. 2016;32(1):25-37.

47. Zilhao R, Papadopoulou B, Courvalin P. Occurrence of the Campylobacter resistance gene tetO in Enterococcus and Streptococcus spp. Antimicrob Agents Chemother. 1988;32(12):1793-6.

48. Kazimierczak KA, Rincon MT, Patterson AJ, Martin JC, Young P, Flint HJ, et al. A new tetracycline efflux gene, tet(40), is located in tandem with tet(O/32/O) in a human gut firmicute bacterium and in metagenomic library clones. Antimicrob Agents Chemother. 2008;52(11):4001-9.

49. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. PLOS ONE. 2015;10(5):e0128036.

50. Schmieder R, Edwards R. Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. PLoS One. 2011;6(3):e17288.

51. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.

52. Kaminski J, Gibson MK, Franzosa EA, Segata N, Dantas G, Huttenhower C. High-Specificity Targeted Functional Profiling in Microbial Communities with ShortBRED. PLOS Computational Biology. 2015;11(12):e1004557.

53. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nature Methods. 2015;12(10):902-3.

54. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nature Methods. 2012;9:811.

55. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. Nature Methods. 2018;15(11):962-8.

56. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable Association Discovery in Population-scale Meta-omics Studies. bioRxiv. 2021:2021.01.20.427420.

57. Shields-Cutler RR, Al-Ghalith GA, Yassour M, Knights D. SplinectomeR Enables Group Comparisons in Longitudinal Microbiome Studies. Frontiers in Microbiology. 2018;9(785).

58. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):R60-R.

59. Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ. 2015;3:e1029-e.

60. Chen EZ, Li H. A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. Bioinformatics. 2016;32(17):2611-7.

61. Murtagh F, Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? Journal of Classification. 2014;31(3):274-95.

62. Ho J, Tumkaya T, Aryal S, Choi H, Claridge-Chang A. Moving beyond P values: data analysis with estimation graphics. Nature Methods. 2019;16(7):565-6.

63. de Lastours V, Maugy E, Mathy V, Chau F, Rossi B, Guérin F, et al. Ecological impact of ciprofloxacin on commensal enterococci in healthy volunteers. Journal of Antimicrobial Chemotherapy. 2017;72(6):1574-80.

64. Edwards R. Resistance to β-Lactam Antibiotics in Bacteroides Spp. Journal of Medical Microbiology. 1997;46(12):979-86.

65. Ferreira LQ, Avelar KES, Vieira JMBD, de Paula GR, Colombo APV, Domingues RMCP, et al. Association Between the cfxA Gene and Transposon Tn4555 in Bacteroides distasonis Strains and Other Bacteroides Species. Current Microbiology. 2007;54(5):348-53.

66. García N, Gutiérrez G, Lorenzo M, García JE, Píriz S, Quesada A. Genetic determinants for cfxA expression in Bacteroides strains isolated from human infections. Journal of Antimicrobial Chemotherapy. 2008;62(5):942-7.

67. Raymond F, Ouameur AA, Déraspe M, Iqbal N, Gingras H, Dridi B, et al. The initial state of the human gut microbiome determines its reshaping by antibiotics. ISME J. 2016;10(3):707-20.

68. Li R, Wang H, Shi Q, Wang N, Zhang Z, Xiong C, et al. Effects of oral florfenicol and azithromycin on gut microbiota and adipogenesis in mice. PLoS One. 2017;12(7):e0181690-e.

69. Korpela K, Salonen A, Virta LJ, Kekkonen RA, Forslund K, Bork P, et al. Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. Nat Commun. 2016;7:10410-.

70. Drew RH, Gallis HA. Azithromycin—Spectrum of Activity, Pharmacokinetics, and Clinical Applications. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy. 1992;12(3):161-73.

71. Borin MT. A Review of the Pharmacokinetics of Cefpodoxime Proxetil. Drugs. 1991;42(3):13-21.

72. Fish DN, Chow AT. The Clinical Pharmacokinetics of Levofloxacin. Clinical Pharmacokinetics. 1997;32(2):101-19.

73. Woodmansey EJ. Intestinal bacteria and ageing. Journal of Applied Microbiology. 2007;102(5):1178-86.

74. Shimizu K, Ogura H, Hamasaki T, Goto M, Tasaki O, Asahara T, et al. Altered gut flora are associated with septic complications and death in critically ill patients with systemic inflammatory response syndrome. Dig Dis Sci. 2011;56(4):1171-7.

75. Wilmanski T, Diener C, Rappaport N, Patwardhan S, Wiedrick J, Lapidus J, et al. Gut microbiome pattern reflects healthy ageing and predicts survival in humans. Nat Metab. 2021;3(2):274-86.

76. Jernberg C, Löfmark S, Edlund C, Jansson JK. Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. ISME J. 2007;1(1):56-66.

77. Moya A, Ferrer M. Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. Trends in Microbiology. 2016;24(5):402-13.

78. Galloway-Peña JR, Smith DP, Sahasrabhojane P, Wadsworth WD, Fellman BM, Ajami NJ, et al. Characterization of oral and gut microbiome temporal variability in hospitalized cancer patients. Genome Medicine. 2017;9(1):21.

79. Milani C, Lugli GA, Duranti S, Turroni F, Bottacini F, Mangifesta M, et al. Genomic encyclopedia of type strains of the genus Bifidobacterium. Appl Environ Microbiol. 2014;80(20):6290-302.

80. Hale I, O'Neill PM, Berry NG, Odom A, Sharma R. The MEP pathway and the development of inhibitors as potential anti-infective agents. MedChemComm. 2012;3(4):418-33.

81. Zinglé C, Kuntz L, Tritsch D, Grosdemange-Billiard C, Rohmer M. Isoprenoid Biosynthesis via the Methylerythritol Phosphate Pathway: Structural Variations around Phosphonate Anchor and Spacer of Fosmidomycin, a Potent Inhibitor of Deoxyxylulose Phosphate Reductoisomerase. The Journal of Organic Chemistry. 2010;75(10):3203-7.

82. Eberl M, Hintz M, Reichenberg A, Kollas A-K, Wiesner J, Jomaa H. Microbial isoprenoid biosynthesis and human γδ T cell activation. FEBS Letters. 2003;544(1-3):4-10.

83. Jacoby GA. AmpC beta-lactamases. Clin Microbiol Rev. 2009;22(1):161-82.

84. Pitout JDD. Multiresistant Enterobacteriaceae: new threat of an old problem. Expert Review of Anti-infective Therapy. 2008;6(5):657-69.

85. Forsberg KJ, Patel S, Wencewicz TA, Dantas G. The Tetracycline Destructases: A Novel Family of Tetracycline-Inactivating Enzymes. Chem Biol. 2015;22(7):888-97.

86. Choy A, Freedberg DE. Impact of microbiome-based interventions on gastrointestinal pathogen colonization in the intensive care unit. Therap Adv Gastroenterol. 2020;13:1756284820939447-.

87. Zhou W, Sailani MR, Contrepois K, Zhou Y, Ahadi S, Leopold SR, et al. Longitudinal multi-omics of host–microbe dynamics in prediabetes. Nature. 2019;569(7758):663-71.

88. Palm NW, de Zoete MR, Cullen TW, Barry NA, Stefanowski J, Hao L, et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. Cell. 2014;158(5):1000-10.

89. Zhao Y, Wu J, Li JV, Zhou N-Y, Tang H, Wang Y. Gut Microbiota Composition Modifies Fecal Metabolic Profiles in Mice. Journal of Proteome Research. 2013;12(6):2987-99.

90. Anthony WE, Burnham C-AD, Dantas G, Kwon JH. The Gut Microbiome as a Reservoir for Antimicrobial Resistance. J Infect Dis. 2020.

91. Leggett RM, Alcon-Giner C, Heavens D, Caim S, Brook TC, Kujawska M, et al. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. Nature Microbiology. 2020;5(3):430-42.

92. Yassour M, Vatanen T, Siljander H, Hämäläinen A-M, Härkönen T, Ryhänen SJ, et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Science Translational Medicine. 2016;8(343):343ra81.

93. Jia W, Li H, Zhao L, Nicholson JK. Gut microbiota: a potential new territory for drug targeting. Nature Reviews Drug Discovery. 2008;7(2):123-9.

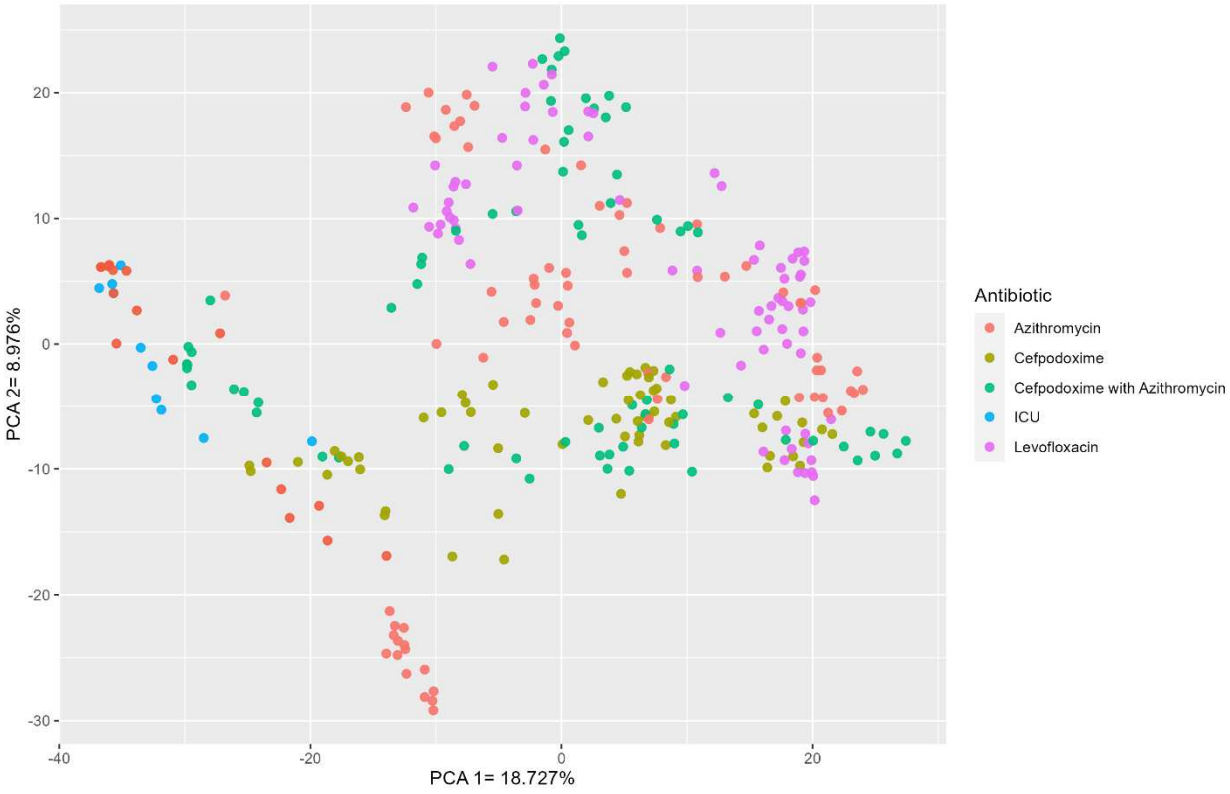
94. Andremont A, Cervesi J, Bandinelli P-A, Vitry F, de Gunzburg J. Spare and repair the gut microbiota from antibiotic-induced dysbiosis: state-of-the-art. Drug Discovery Today. 2021.

95. Anthony WE, Wang B, Sukhum KV, D'Souza AW, Hink T, Cass C, et al. Acute and persistent effects of commonly used antibiotics on the gut microbiome and resistome in healthy adults. Cell Reports. 2022;39(2).

2.7 Appendix (Supplemental Material)

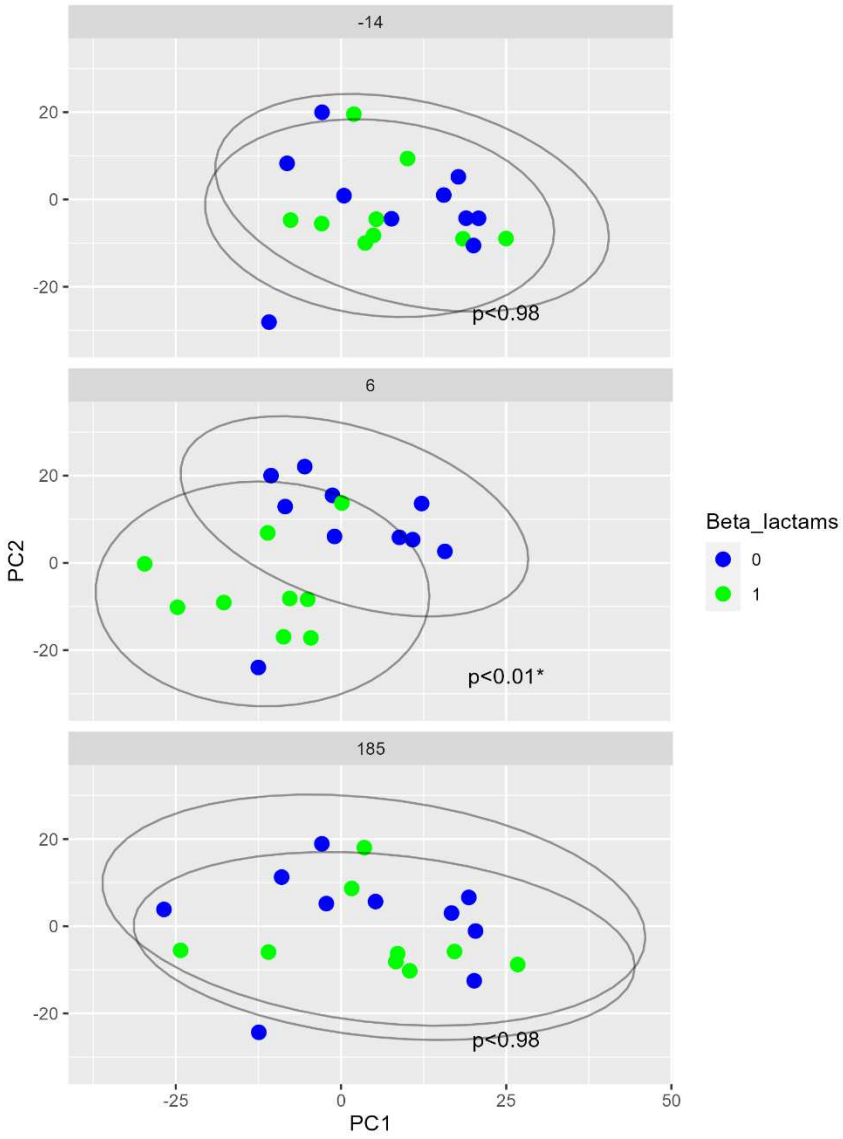
2.7.1 Figures

Figure 2. 8S **PCA of the metagenomic samples from healthy volunteers and ICU cohort** (Supplementary Figure S1)



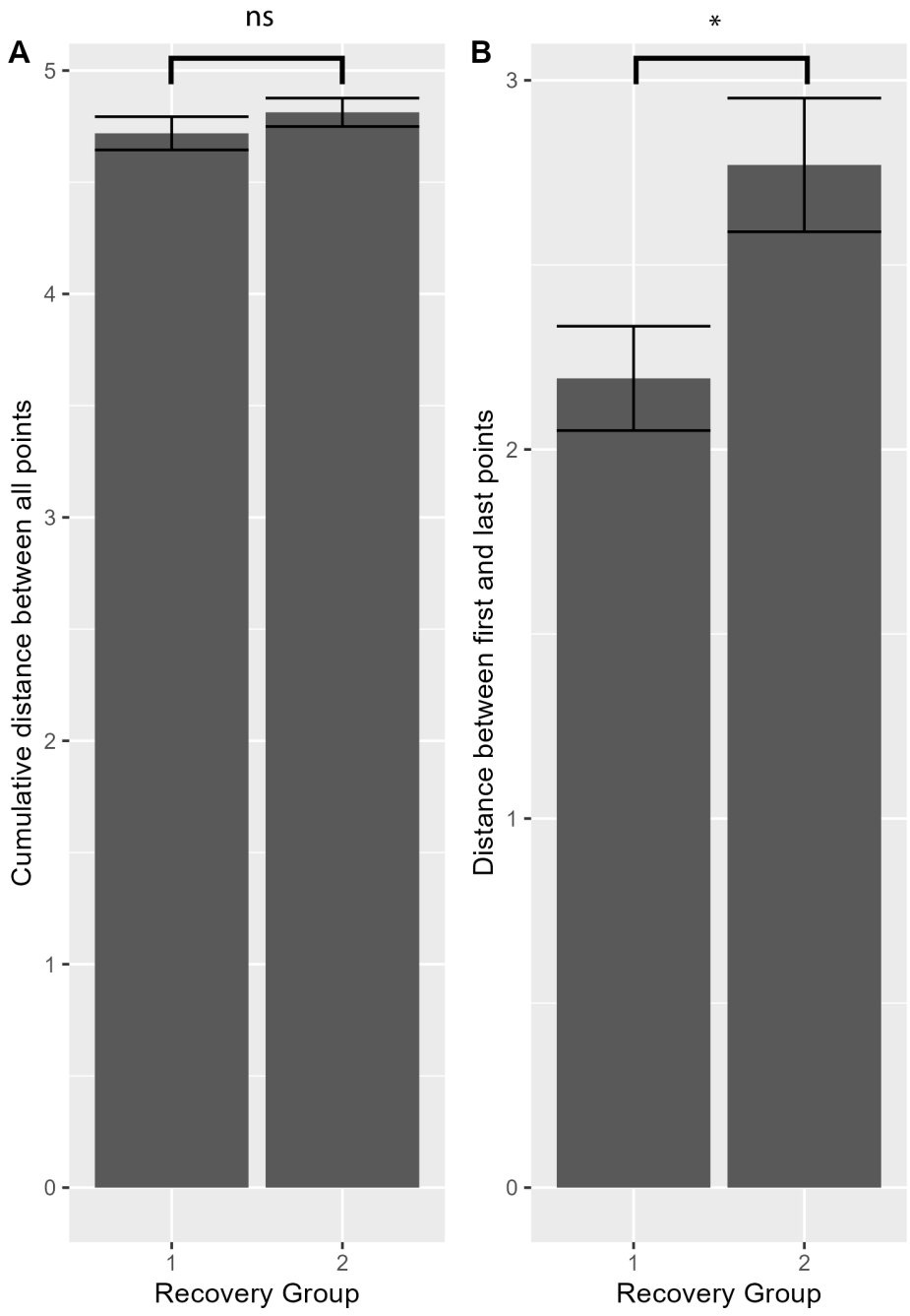
Supplementary Figure S1. PCA of the metagenomic samples from healthy volunteers and ICU cohort. Related to figures 1 and 5. Samples are colored by which antibiotic treatment group they were a part of, or whether they were samples from the ICU. Percentages on axes labels represent the amount of variation in the dataset explained by that axis.

Figure 2. 9S **PCA of the antibiotic perturbance in metagenomic samples from healthy volunteers** (Supplementary Figure S2)



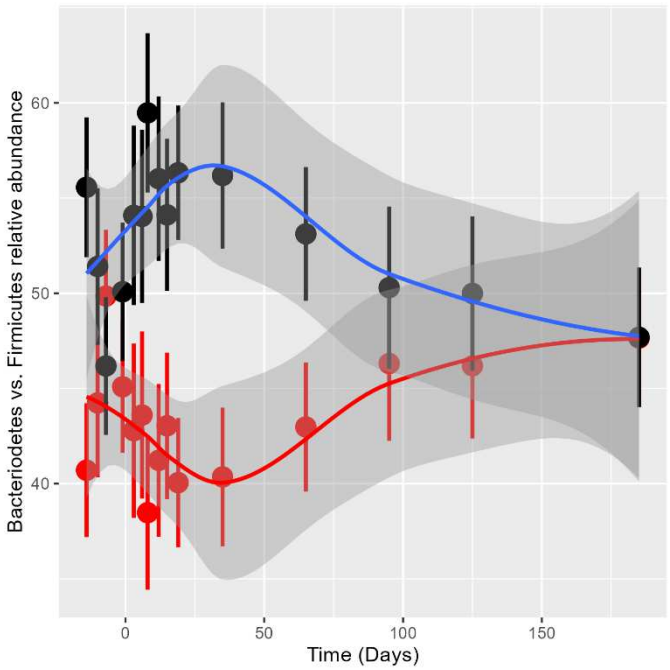
PCA of the antibiotic perturbance in metagenomic samples from healthy volunteers. Related to figure 1. Samples are colored green to denote they were from either the CPD or CPD+AZM volunteers, and blue to denote they were from the LVX and AZM conditions. P values are from PERMANOVA tests between groups, and grey circles represent 95% CI ellipses.

*Figure 2. 10***S Distance between first and last sample versus cumulative traveled distance in PCA** (Supplementary Figure S3)



Distance between first and last sample versus cumulative traveled distance in PCA. Related to figure 2. A. The difference in the cumulative distance traveled between slow and fast recovery groups after antibiotic treatment. NS means non-significant difference. B. The distance between first and last points of a volunteer's microbiome samples in the FRG and SRG. * refers to a difference with a p value < 0.05 between the two groups. Tests were conducted via Student's T-test on square root transformed distances.

*Figure 2. 1*1S **Delay in the increase in relative abundance of firmicutes after antibiotic treatment.** (Supplementary Figure S4)



Delay in the increase in relative abundance of *Firmicutes* after antibiotic treatment. Related to figures 2 and 3. Dots represent mean relative abundance of *Firmicutes* (red) versus *Bacteriodetes* (black) with bootstrapped confidence intervals. A loess regression was fit to these data (blue and red lines). *Firmicutes* mean relative abundance begins to increase at day 65, while *Bacteriodetes* decreases.

2.7.2 Tables

*Table 2. 1*S **Demographics Table (Table S1).** Related to STAR methods.

Characteristics	**N (%)**
Female	10 (50)
Race	
White	17 (85)
African-American	2 (10)
Asian / Pacific Islander	1 (5)
Hispanic	1 (5)
Age (median [range])	37 (24 – 59)

Healthcare worker	6 (30)
BMI (median [range])	24 (18.5 – 51)
Normal	13 (65)
Overweight	4 (20)
Obese	3 (15)
Special diet*	5 (25)
Vegetarian	2 (10)
Shrimp allergy	1 (5)
Gluten free	1 (5)
Not specified	1 (5)
Pre-existing medical comorbidity	3 (15)
Hyperlipidemia	1 (5)
Glomerular thin basement membrane disease	1 (5)
Benign prostate hyperplasia and allergies	1 (5)

*Vegetarian (2), shrimp allergy (1), gluten free (1), not specified (1)

**Hyperlipidemia (1), Glomerular thin basement membrane disease (1), benign prostate hyperplasia and allergies (1)

*Table 2. 2***S** **Sampling Times (Table S2).** Related to STAR methods.

Specimen Date	**Samples (n)**	**Cultured**
Day 14 pre-ABX	20	yes
Day 10 pre-ABX	20	no

Day 7 pre-ABX	20	no
Day 1 pre-ABX	19	yes
Day 3	19	yes
Day 6	19	yes
Day 8	19	yes
Day 12	20	yes
Day 15	19	no
Day 19	19	yes
Day 35	19	yes
Day 65	20	yes
Day 95	17	no
Day 125	19	no
Day 185	20	yes

*Table 2. 3*S **Aerobic CFU change over time in healthy volunteer microbiomes after antibiotic perturbation (Table S3).** Related to STAR methods.

.y.	group1	group2	p.adj (BY)	p.signif	method

logCFU	**-14**	**-1**	**1**	**ns**	**Paired Wilxocon**
logCFU	**-14**	**3**	**0.57**	**ns**	**Paired Wilxocon**
logCFU	**-14**	**6**	**0.0074**	******	**Paired Wilxocon**
logCFU	**-14**	**8**	**0.26**	**ns**	**Paired Wilxocon**
logCFU	**-14**	**12**	**0.0074**	******	**Paired Wilxocon**
logCFU	**-14**	**19**	**1**	**ns**	**Paired Wilxocon**
logCFU	**-14**	**35**	**1**	**ns**	**Paired Wilxocon**
logCFU	**-14**	**65**	**1**	**ns**	**Paired Wilxocon**
logCFU	**-14**	**185**	**0.26**	**ns**	**Paired Wilxocon**

*Table 2. 4***S** **Anaerobic CFU change over time in healthy volunteer microbiomes after antibiotic perturbation (Table S4).** Related to STAR methods.

.y.	**group1**	**group2**	**p.adj (BY)**	**p.signif**	**method**
logCFU	**-14**	**-1**	**1**	**ns**	**Paired Wilxocon**
logCFU	**-14**	**3**	**1**	**ns**	**Paired Wilxocon**
logCFU	**-14**	**6**	**0.0047**	******	**Paired Wilxocon**
logCFU	**-14**	**8**	**1**	**ns**	**Paired Wilxocon**

logCFU	**-14**	**12**	**0.0047**	******	**Paired Wilxocon**
logCFU	**-14**	**19**	**1**	**ns**	**Paired Wilxocon**
logCFU	**-14**	**35**	**1**	**ns**	**Paired Wilxocon**
logCFU	**-14**	**65**	**1**	**ns**	**Paired Wilxocon**
logCFU	**-14**	**185**	**0.095**	**ns**	**Paired Wilxocon**

*Table 2. 5*S **Metagenomic species change over time in healthy volunteer microbiomes after antibiotic perturbation (Table S5).** Related to STAR methods.

.y.	group1	group2	p.adj (BY)	p.signif	method
Specnum	**-14**	**-10**	**1**	**ns**	**Paired Wilxocon**
Specnum	**-14**	**-7**	**1**	**ns**	**Paired Wilxocon**
Specnum	**-14**	**-1**	**0.076**	**ns**	**Paired Wilxocon**
Specnum	**-14**	**3**	**0.29**	**ns**	**Paired Wilxocon**
Specnum	**-14**	**6**	**0.0052**	******	**Paired Wilxocon**
Specnum	**-14**	**8**	**0.0052**	******	**Paired Wilxocon**
Specnum	**-14**	**12**	**0.0144**	*****	**Paired Wilxocon**

Specnum	**-14**	**15**	**0.032**	*****	**Paired Wilxocon**
Specnum	**-14**	**19**	**0.21**	**ns**	**Paired Wilxocon**
Specnum	**-14**	**35**	**0.19**	**ns**	**Paired Wilxocon**
Specnum	**-14**	**65**	**0.76**	**ns**	**Paired Wilxocon**
Specnum	**-14**	**95**	**1**	**ns**	**Paired Wilxocon**
Specnum	**-14**	**125**	**0.51**	**ns**	**Paired Wilxocon**
Specnum	**-14**	**185**	**0.76**	**ns**	**Paired Wilxocon**

*Table 2. 6***S** **Resistance change over time in CPD, CPD+AZM, and AZM treated healthy volunteer microbiomes after antibiotic perturbation (Table S6).** Related to STAR methods.

.y.	**group1**	**group2**	**p.adj (BY)**	**p.signif**	**method**
ResNormBurden	**Pre-average**	**3**	**1**	**ns**	**Paired Wilxocon**
ResNormBurden	**Pre-average**	**6**	**0.003**	******	**Paired Wilxocon**
ResNormBurden	**Pre-average**	**8**	**0.028**	*****	**Paired Wilxocon**
ResNormBurden	**Pre-average**	**12**	**0.003**	******	**Paired Wilxocon**
ResNormBurden	**Pre-average**	**15**	**0.003**	******	**Paired Wilxocon**
ResNormBurden	**Pre-average**	**19**	**0.002**	******	**Paired Wilxocon**

ResNormBurden	**Pre-average**	**35**	**0.002**	**	**Paired Wilxocon**
ResNormBurden	**Pre-average**	**65**	**0.019**	*	**Paired Wilxocon**
ResNormBurden	**Pre-average**	**95**	**0.005**	**	**Paired Wilxocon**
ResNormBurden	**Pre-average**	**125**	**0.01**	*	**Paired Wilxocon**
ResNormBurdens	**-14**	**185**	**0.003**	**	**Paired Wilxocon**

*Table 2. 7***S** **Biomarker identification table from metaphlan lefse results (Table S7).** Related to STAR methods.

A: Clostridiaceae
B: Clostridium
C: Lachnospiraceae
D: Coprococcus
E: Blautia
F: Lachnospiraceae noname
G: Anaerostipes
H: Roseburia
I: Ruminococcus
J: Eubacteriaceae
K: Eubacterium
L: Clostridiales noname
M: Clostridiales noname
N: Erysipelotrichaceae
O: Coprobacillus
P: Bacteroidaceae
Q: Bacteroides
R: Coriobacteriaceae
S: Collinsella

*Table 2. 8***S** **Biomarker identification table from shortbred lefse results (Table S8).** Related to STAR methods.

A: TEM 9Escherichia coli
AA: NP 414996 1
AB: CREscherichia coli
AC: BAE77933 1
AD: emrEscherichia coli
AE: BAA11236 1
AF: gadEscherichia coli
AG: ANK04027 1
AH: emrEscherichia coli
AI: NP 417169 1
AJ: mdtEscherichia coli
AK: BAE77781 1
AL: leuEscherichia coli
AM: NP 414618 4
AN: acrEscherichia coli
AO: NP 414995 1
AP: tolEscherichia coli
AQ: ACN32294 1
AR: gadEscherichia coli
AS: YP 491919 1
AT: gadEscherichia coli
AU: NP 417969 1
AV: emrEscherichia coli
AW: BAA11237 1
AX: acrEscherichia coli
AY: AAC76297 1
AZ: MFS Antibiotic Efflux Pump
Aa: emrEscherichia coli
Ab: EFF04178 1
Ac: mdtEscherichia coli
Ad: AAC74149 2
Ae: mdtEscherichia coli
Af: BAE78083 1
Ag: mdtEscherichia coli
Ah: AAC76733 1
Ai: mdtEscherichia coli
Aj: AAC75138 1
Ak: mdtEscherichia coli
Al: AAC77293 1
Am: mdtEscherichia coli
An: YP 489321 1
Ao: mdtEscherichia coli
Ap: BAE78082 1
Aq: efmEnterococcus faecium
Ar: BAG75524 1
As: mdtEscherichia coli
At: BAE78084 1
Au: ABC Antibiotic Efflux Pump
Av: msbEscherichia coli
Aw: NP 415434 1
Ax: MATE Antibiotic Efflux
Ay: SMR Antibiotic Efflux
Az: Escherichia coli emrE
Ba: AAK30619 1
BA: CAA77936 1
BB: Trimethoprim Resistant Dihydrofolate Reductase dfr
BC: dfrEnterococcus faecalis
BD: AAD01868 1
BE: Tetracycline Inactivation Enzyme
BF: tetBacteroides fragilis
BG: AAA27471 1
BH: tetBifidobacterium longum
BI: ACA23185 1
BJ: Tetracycline Resistant Ribosomal Protection Protein
BK: tet3Clostridiaceae bacterium
BL: CAC41371 1
BM: tetClostridium difficile
BN: CAI67339 1
BO: tetBacteroides fragilis
BP: CAA79727 1
BQ: tetCampylobacter jejuni
BR: AAA23033 1
BS: MFS Antibiotic Efflux Pump
BT: tet 4 Streptococcus suis
BU: AFK31666 1
BV: tet Geobacillus stearothermophilus
BW: AAA22851 1
BX: Gene Altering Cell Wall Charge
BY: PmrEscherichia coli
BZ: AAC75089 1
Ba: PmrEscherichia coli
Bb: AAC75314 1
Bc: PmrEscherichia coli
Bd: BAE78116 1
Be: Other EffluComplex Subunit
Bf: Aminoglycoside NucleotidyltransferasANT
Bg: ANT ICampylobacter jejuni
Bh: AJ27543 1
Bi: Other EffluComplex Subunit
Bj: RND Antibiotic Efflux
Bk: acrEscherichia coli
Bl: YP 490697 1
Bm: Aminoglycoside AcetyltransferasAAC
Bn: AAC IEnterococcus faecium
Bo: AAB65333 1
Bp: AAC Ie APH IStaphylococcus aureus
Bq: NP 863643 1
Br: Glycopeptide Resistance Gene Cluster
Bs: vanSEnterococcus faecium
Bt: AAA65954 1
Bu: vanHEnterococcus faecium
Bv: AAA65955 1
Bw: vanEnterococcus faecium
Bx: AAA65956 1
By: vanXEnterococcus faecium
Bz: AAA65957 1
C: TEM 19Shigella sonnei
CA: vanZEnterococcus faecium
CB: AAA65959 1
CC: vanYEnterococcus faecium
CD: AAA65958 1
CE: vanREnterococcus faecium
CF: AAA65953 1
CG: Other EffluComplex Subunit
CH: ABC Antibiotic Efflux Pump
CI: msrEnterococcus faecium
CJ: AAK01167 1
CK: msrEscherichia coli
CL: NP 416292 1
CM: 23S ribosomal RNA methyltransferase
CN: ErmEnterococcus faecium
CO: AAF86219 1
CP: ErmLysinibacillus sphaericus
CQ: AAA22419 1
CR: ErmBacteroides thetaiotaomicron
CS: AAC37034 1
CT: Lincosamide NucleotidyltransferasLIN
CU: InuStreptococcus agalactiae
CV: AAY32951 1
CW: MFS Antibiotic Efflux Pump
CX: mefExiguobacterium sp
CY: YP 008997285 1
CZ: Quinolone Resistance Proteiqnr
Ca: mfEscherichia coli
Cb: NP 415632 1
Cc: Other EffluComplex Subunit
Cd: emrEscherichia coli
Ce: BAA16547 1
Cf: emrEscherichia coli
Cg: AAC75733 1
Ch: palEscherichia coli
Ci: NP 417544 5
Cj: Gene Conferring Resistance via Molecular Bypass
Ck: bacEscherichia coli
Cl: AAC76093 1
Cm: ABC Antibiotic Efflux Pump
D: AFE48832 1
E: TEM 5Escherichia coli
F: ACJ43254 1
G: TEM 14Enterobacter cloacae
H: AAX56615 1
I: TEM 14Escherichia coli
J: AAZ14084 2
K: TEM 9Proteus mirabilis
L: AAF66653 1
M: TEM 13Salmonella enterica
N: AAW47922 1
O: CfxAParabacteroides distasonis
P: AAV37206 1
Q: CblA mixed culture
R: ACT97415 1
S: TEM 7Escherichia coli
T: AAP05611 1
U: CfxAPrevotella intermedia
V: AAD23513 1
W: TEM 5Klebsiella oxytoca
X: AAD45935 1
Y: TEM 6Providencia stuartii
Z: AAC05975 1
a: TEM 1Klebsiella oxytoca
b: AAA25033 1
c: TEM 10Salmonella enterica
d: AAM28884 1
e: CfxBacteroides vulgatus
f: AAB17891 1
g: Other EffluComplex Subunit
h: marEscherichia coli
i: YP 489794 1
j: acrEscherichia coli
k: AAC76298 1
l: mdtEscherichia coli
m: AAC76539 1
n: baeEscherichia coli
o: YP 490321 1
p: mdtEscherichia coli
q: AAC75135 2
r: evgEscherichia coli
s: NP 311275 1
t: acrEscherichia coli
u: AAC76296 1
v: H NEscherichia coli
w: NP 309766 1
x: cpxEschorichia coli
y: NP 312864 1
z: Escherichia coli acrA

Chapter 3:

Engineering Diverse Fatty Acid Compositions of Phospholipids in Escherichia coli

3.1 Introduction

3.1.1 Abstract

Bacterial fatty acids (FAs) are an essential component of the cellular membrane and are an important source of renewable chemicals as they can be converted to fatty alcohols, esters, ketones, and alkanes, and used as biofuels, detergents, lubricants, and commodity chemicals. Most prior FA bioconversions have been performed on the carboxylic acid group. Modification of the FA hydrocarbon chain could substantially expand the structural and functional diversity of FA-derived products. Additionally, the effects of such modified FAs on the growth and metabolic state of their producing cells are not well understood. Here we engineer novel *Escherichia coli* phospholipid biosynthetic pathways, creating strains with distinct FA profiles enriched in ω7-unsaturated FAs (ω7-UFAs, 75%), Δ5-unsaturated FAs (Δ5-UFAs, 60%), cyclopropane FAs (CFAs, 55%), internally-branched FAs (IBFAs, 40%), and Δ5,ω7-double unsaturated FAs (DUFAs, 46%). Although bearing drastically different FA profiles in phospholipids, UFA, CFA, and IBFA enriched strains display wild-type-like phenotypic profiling and growth. Transcriptomic analysis reveals DUFA production drives increased differential expression and the induction of the *fur* iron starvation transcriptional cascade, but higher TCA cycle activation compared to the UFA producing strain. This likely reflects a slight cost imparted for DUFA production, which resulted

in lower maximum growth in some, but not all, environmental conditions. The IBFA-enriched strain was further engineered to produce free IBFAs, releasing 96 mg/L free IBFAs from 154 mg/L of the total cellular IBFA pool. This work has resulted in significantly altered FA profiles of membrane lipids in *E. coli*, greatly increasing our understanding of the effects of FA structure diversity on the transcriptome, growth, and ability to react to stress.

3.1.2 Introduction

A bacterial cell is a tightly controlled, semi-closed system which is constantly reacting to the effects of an ever-changing environment. The cellular membrane is the first line of cellular defense, compartmentalizing the biochemical processes necessary for cell survival(1, 2). Phospholipids are a major component of cell membranes and play key roles in cell growth, transport, metabolism, survival, and stress tolerance(1, 3, 4). In Gram-negative bacteria, the fatty acid (FA) profile of lipid membranes mostly consists of straight-chain saturated FAs, with some unsaturated and cyclopropane FAs, while in Gram-positive bacteria, varying amount of terminally branched FAs are observed(5, 6).

Due to their efficient biosynthesis, microbial FAs have been recognized as important intermediates for the renewable production of biofuels, commodity chemicals, detergents, lubricants, and polymer precursors(7-14). Unfortunately, natural microbial FAs exhibit high melting temperatures, making their derived products suffer from undesirable properties, such as for bacterial FA-derived biofuels(11, 15, 16). Bacteria exhibit the ability to adapt to different environments and growth stages by altering the composition of their phospholipid FAs(17, 18). *Escherichia coli ML30*, for example, increases the proportion of unsaturated FAs from 47% to 62% when the temperature drops from 37°C to 10°C (3). Modulating membrane composition is an established bioengineering concept called membrane engineering and has even been used to

increase *E. coli* tolerance to bioproducts such as carboxylic acids, alcohols, and aromatic compounds, as well as adverse conditions, such as low temperature(19, 20). Expanding the diversity of bacterial produced FA-derived products would increase their applicability and value but necessitates methods to predictably modify the chemical structure of FA species.

Previous work on engineering free FA (FFA) pathways focused on modifying the carboxylic acid group(12, 21-23) and the ω-terminal carbon to other functional groups(24-26), or modulating the chain-length of FFAs(27). In comparison, altering the internal FA hydrocarbon chain can provide a new class of renewable compounds for various applications, but this has been seldom explored. This is potentially because enzymes responsible for modifying internal hydrocarbon chains only use phospholipid FAs as substrates(1, 28), and modifying the membrane phospholipid can have significant detrimental effects to cell growth and viability under stressful conditions(29). For example, lipid incorporation of medium-chain FAs that are shorter than native FAs can lead to membrane damage and drastically reduce cell viability(30, 31). Similarly, incorporation of exogenously-fed, non-native polyunsaturated FAs into membrane phospholipids can change expression of genes related to cellular respiration, membrane integrity, and oxidative stress(32). Increasing unsaturated FAs (UFAs) in *E. coli* has been attempted before, however the limited structural diversity of naturally produced UFAs has relegated these efforts to monounsaturated FAs and cyclopropane FAs (CFAs)(4, 33, 34).

In this study, we introduce engineered phospholipid biosynthetic pathways into the *E. coli* chassis to produce bacterial strains with greatly diversified, yet highly controllable FA profiles, resulting in FA compositions substantially different from that of the native cell. These strains are enriched in ω7-unsaturated fatty acids (ω7-UFAs; C16:1,Δ9 and C18:1, Δ11), cyclopropane fatty acids (CFAs; C17:0, cyclo9 and C19:0, cyclo11), Δ5-monounsaturated fatty acids (Δ5-MUFAs;

C16:1, Δ5), doubly unsaturated fatty acids (DUFAs; C16:2, Δ5Δ9 and C18:2, Δ5Δ11), or internally-branched fatty acids (IBFAs; C17:0 Me10 and C19:0 Me12), with later three types of FAs non-native to *E. coli* (Figure 3.1).

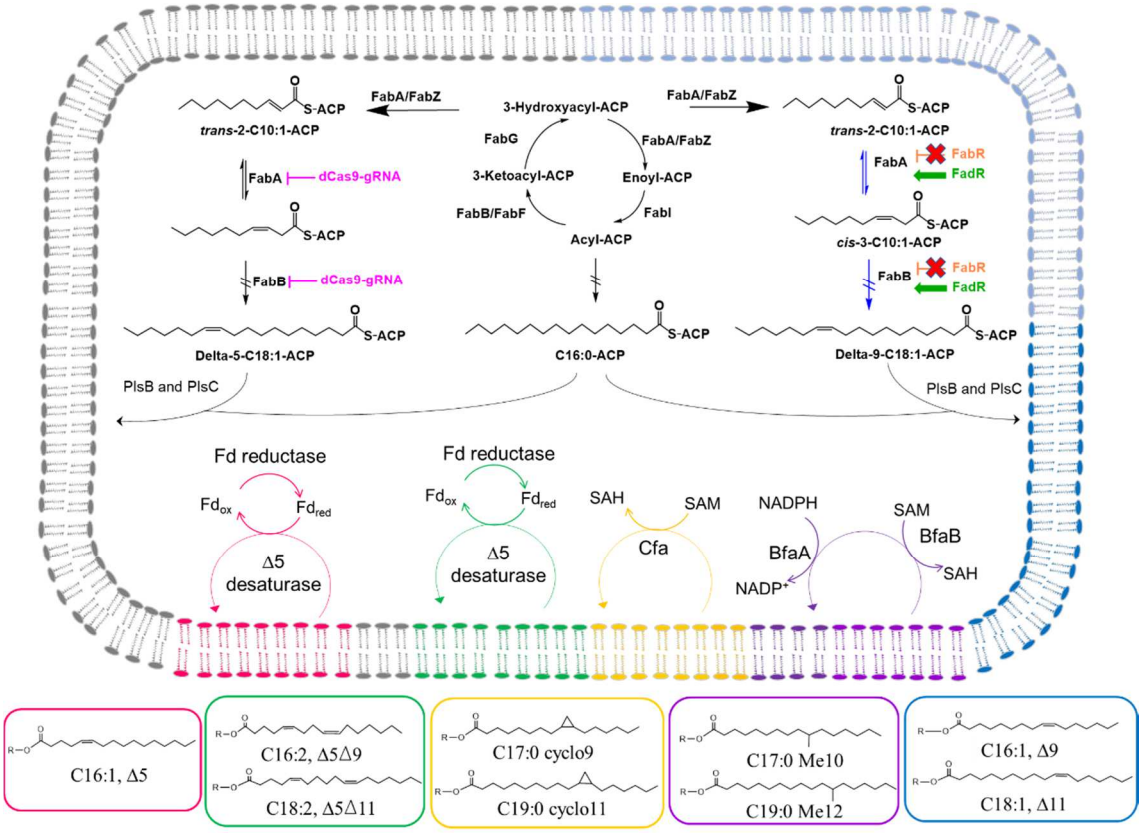


Figure 3. 3 Engineered pathways for the biosynthesis of different FAs in membrane phospholipids in E. coli. Each type of FAs is boxed in the same colored as their biosynthetic pathways. FabA or FabZ: β□hydroxyacyl-ACP dehydrase; FabB: β-ketoacyl-acyl ACP synthase □; FabF: β-ketoacyl-acyl ACP synthase □; FabG: β□ketoreductase; FabH: β-ketoacyl-acyl carrier protein (ACP) synthase □; FabI or FabK: enoyl-ACP reductase; PlsB: glycerol-3-phosphate acyltransferase; PlsC: 1□acyl□sn□glycerol□3-phosphate acyltransferase; Fd: ferredoxin; SAM: S-adenosyl-L-methionine; SAH: S-adenosyl-L-homocysteine; BfaA: NADPH dependent oxidoreductase; BfaB: SAM-dependent methyltransferase; Cfa: cyclopropane fatty acid synthase.

The produced DUFAs and IBFAs (structurally different from tuberclostearic acid, C19:0 Me10) are extremely rare in nature and have only been identified in trace amounts from the seed of the *Ephedra* plant and from sulfide-forming bacteria, respectively(35). We further investigate the effect of modifying cell FA composition on bacterial growth, viability in stressful conditions, and

on the metabolic networks of the cell. CFA- and IBFA-enriched strains exhibited similar growth rates, cell densities, and responses to environmental stress as wild-type *E. coli* under a wide range of conditions. DUFA enrichment resulted in reduced viability across multiple environmental conditions, a markedly different transcriptomic profile, as well as lowered cell densities in glycerol and glucose supplemented with high salt. These designed strains substantially diversify phospholipid FA composition within *E. coli*, and through the demonstrated production of free IBFAs (FIBFAs) create new opportunities in biotechnology as microbial hosts for chemical production(19).

3.2 Results

3.2.1 Engineering E. coli to modulate phospholipid profile

Wild-type *E. coli* K-12 strains contain 30-60% UFAs with a *cis* double bond located precisely at the ω7 position (ω7-UFAs)(6, 36, 37). ω7-UFAs can be used as intermediates for synthesizing other FA structures such as CFA and internally branch-chain fatty acids (IBFAs)(6, 38, 39). Accordingly, we started by engineering *E. coli* to enrich ω7-UFAs in phospholipids. The FA degradation pathway was first deactivated by deleting *fadE*, whose enzyme product catalyzes the first step in β-oxidation(13). To create a strain that produced only ω7-UFAs in addition to saturated straight-chain FAs, the CFA synthase (encoded by *cfa*) that converts ω7-UFA to CFA was deleted from the *E. coli* genome. The resulting strain, named ω7-UFA-1, significantly increased ω7-UFA composition to 50.0±0.4%, in comparison to 31.5±1.2% (student's t-test, adjusted p value = 0.002) in wild-type *E. coli* (Figure 3.2a). Next, the FA transcriptional regulator FadR, which activates the expression of two ω7-UFA biosynthesis genes *fabA* and *fabB,* was overexpressed, significantly increasing ω7-UFA composition to 55.8±0.0% (Strain ω7-UFA-2) versus WT

(student's t-test, adjusted p value = 0.002). To further enrich ω7-UFA, three strategies were explored in parallel to optimize *fabA* and *fabB* expression. In the first strategy, an additional copy of *fabB* or *fabAB* was overexpressed from a plasmid, which increased ω7-UFA compositions to 75.0±5.2% (student's t-test, adjusted p value = 0.007 versus WT) and 72.0±2.0% (student's t-test, p value = 0.002 versus WT), respectively (Figure 3.2a and Figure 3.8Sa). The third strategy involved the deletion of FabR, another FA transcriptional regulator that represses the expression of *fabA* and *fabB*. Deletion of FabR increased ω7-UFA composition to 69.4±1.0% (student's t test, adjusted p value = 0.002). Of these engineered variants, strain ω7-UFA-5 produced the highest ω7-UFA titer, reaching 197.7±16.3 (adjusted p value < 0.05 versus all strains) mg/L (Figure 3.8Sa). The titer achieved in this work is 5.7-fold higher than previous efforts solely relying on overexpression of *fabAB* alone, which do not prevent *cfa* formation(33).

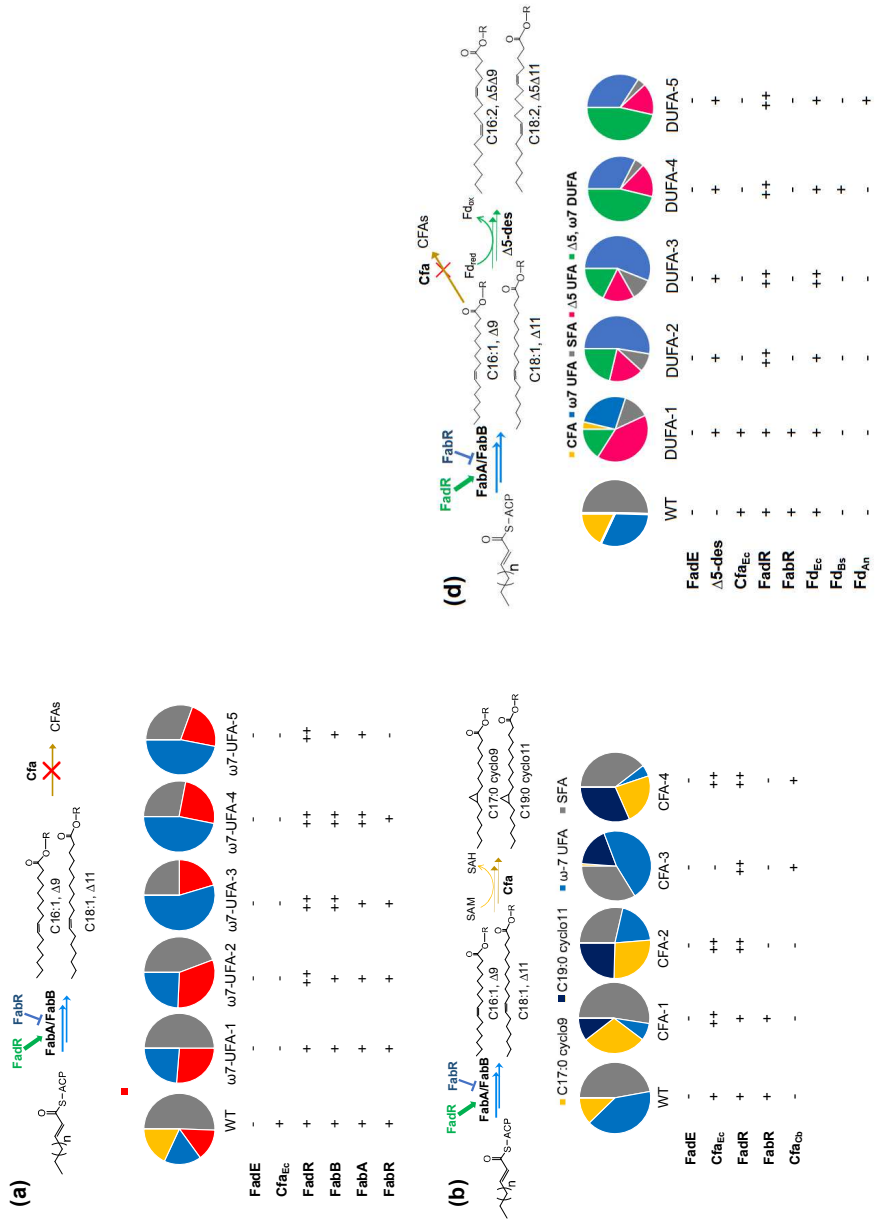


Figure 3. 4 Engineering E. coli to produce ω7-unsaturated fatty acids. (ω7-UFAs, a). cyclopropane fatty acids (CFAs, b). Internally branched fatty acids (IBFAs, c). doubly unsaturated fatty acids (DUFAs, d). and Δ5-unsaturated fatty acids (Δ5-MUFAs, e). in phospholipids. The engineered biosynthetic pathways are shown on the top of each figure. Pie-charts indicate the FA composition in the phospholipid of each strain. Genetic changes for each strain are shown under each Pie-chart. - indicates gene deletion from E. coli's genome; + indicates the native gene is not deleted; ++ indicates gene overexpression from plasmid DNA.

The enhanced titers presented here may also be attributed to the overexpression of *fadR*, which has been shown to enhance overall titers of FAs through a more global tuning of FA pathway genes(40).

CFAs are naturally synthesized in *E. coli* (Figure 3.1) and are important for bacterial resistance to environmental stresses, such as acid and high osmotic pressure(2, 41). However, WT *E. coli* only accumulates between 2-20% of CFAs during logarithmic growth(42, 43), and our wild-type *E. coli* only accumulated 12.5±2.3% of CFA in phospholipid when tested. To enhance CFA composition in phospholipid, we overexpressed CFA synthase in the *fadE*-deleted *E. coli* strain and observed an increase in the proportion of CFAs to 39.7±2.2% (p value = 0.006, Student's t-test versus WT), consisting of 29.2±2.4% C17:0 CFA and 10.5±0.6% C19:0 CFA (strain CFA-1, Figure 3.2b) as identified by Gas Chromatography-Mass Spectrometry (GC-MS) (Figure 3.10Sa-b). CFA is biosynthesized by methylation of ω7-UFA(6). To enhance the cellular pool of ω7-UFA, FabR was deleted and FadR was overexpressed. The resulting strain (CFA-2) produces 53.7±4.5% CFAs and 19.8±4.5% ω7-UFA. *E. coli* Cfa methylates ω7-UFA preferably at the sn-2 position in phospholipids. To convert the remaining ω7-UFA to CFA, the *Clostridium butyricum* Cfa, which prefers the sn-1 position(5), was overexpressed. While overexpression *C. butyricum* Cfa alone produced 19.2±3.9% CFA (strain CFA-3), coexpression of both *E. coli* Cfa and *C. butyricum* Cfa (strain CFA-4) increased CFA composition to 55.3±0.3% (p value = 0.004 Student's t-test versus WT), with a titer of 84.7±4.7 (p value = 0.003 Student's t-test versus WT) mg/L (Figure 3.8Sb).

Internally branched fatty acids (IBFAs) have substantially lower melting temperatures compared to straight-chain FAs of the same length, and are thus attractive precursors for jet-fuels, which demand low freezing points. However, IBFAs are mostly found in *Mycobacterium tuberculosis* and a few other *Mycobacterium* and *Rhodococcus* species(6, 44). To produce IBFAs in *E. coli*, we overexpressed the *R. opacus* PD630 IBFA biosynthetic pathway, which contains a

S-adenosyl-l-methionine-dependent methyltransferase (BfaB) and a FAD-binding oxidoreductase (BfaA)(6, 44). The resulting *E. coli* strain converts palmitoleic acid (C16:1Δ9) and vaccenic acid (C18:1Δ11) to 10-methylhexadecanoic acid (C17:0Me10) and 12-methyl octadecanoic acid (C19:0Me12), respectively, with a total IBFA composition of 11.7% (Strain IBFA-1, average of two replicates, Figure 3.2c, Figure 3.8Sc, 3.9S and 3.8Sc-d). To further increase IBFA composition, Cfa was deleted to avoid competition with the ω7-UFA precursor, and ω7-UFA biosynthesis was enhanced by FabR deletion and FadR overexpression. These engineering strategies led to a gradual increase of IBFA composition up to 30.9±1.2% (p value = 2.0E-04, Students t-test versus IBFA-1). Additionally, BfaAB enzymes from different microorganisms were screened. The *Mycobacterium smegmatis* enzymes (strain IBFA-5) produced the highest IBFA composition, up to an average of 39.2±1.2% with an IBFA titer of 91.3±4.6 mg/L (p value = 7.1E-05, Student's t test versus IBFA-1).

Double unsaturated FAs (DUFAs) are intermediates in the biosynthesis of polyunsaturated FAs that have various health benefits due to their abilities in scavenging radical species(45). To produce DUFAs in *E. coli*, we overexpressed the Δ5-desaturase (encoded by Δ*5-des*) from *Bacillus subtilis* in the *fadE*-deleted *E. coli* strain. The resulting strain DUFA-1 consists of 16.0±0.2% of Δ5,ω7 DUFA (mostly C18:2, Δ5Δ11, Figure 3.10Se-f), 26.3±1.2% ω7-UFA, and 40.9±0.7% Δ5 C16:1 UFA (Figure 3.2d) Increasing the precursor pool of ω7-UFAs by Cfa deletion, FabR knockout, and FadR overexpression increased Δ5,ω7 DUFA to 21.3±1.0% (p value = 0.007, Student's t test versus DUFA-1) (strain DUFA-2), but at the same time resulting in a high ω7-UFA composition (52.6±0.8%), suggesting that the pathway is limited by the Δ5-desaturase. Δ5-desaturase uses a ferredoxin as an electron donor to reduce C-C single bonds at the Δ5 position(46). To optimize Δ5-desaturase activity, the ferredoxin gene from *E. coli*, *B. subtilis*, and *Anabaena* sp.

7120 was individually overexpressed. Ferredoxin from either *B. subtilis* (strain DUFA-4) or *Anabaena* sp. 7120 (strain DUFA-5) increased Δ5,ω7 DUFA composition to 45.4±0.6% (p value = 1.58E-05, Student's t test versus DUFA-1) and 44.8±3.8% (p value = 0.002, Student's t test versus DUFA-1) in phospholipid with a titer of 151.3±3.0 mg/L and 153.9±16.2 mg/L, respectively (Figure 3.2d, Figure 3.8Sd).

While most monounsaturated FAs (MUFAs) contain the double bond at ω7 position, double bonds at alternative positions, such as Δ5, are rare in nature and have only been found in small amounts (< 9%) in *B. subtilis*(35) and in the seeds of *Brassica* species. The unique double bond position may enable region-selective labeling and other chemical modifications on FAs for various applications(47). To produce Δ5-MUFA, the *B. subtilis* Δ5-desaturase (encoded by Δ*5-des*) was overexpressed in the Δ*fadE*-deleted *E. coli* strain. The resulting strain (Δ5-MUFA-1) contains only 7.8±1.7% of Δ5-MUFAs (Figure 3.10Sg), with 10.4±1.1% of ω7-UFAs and 18.6±1.7% CFAs (Figure 3.3c). Deletion of Cfa to eliminate CFAs enhanced Δ5-MUFAs to 39.5±7.4% (p value = 0.0001, Student's t test versus MUFA-1). To further increase the proportion of Δ5-MUFAs, we aimed to decrease ω7-UFA biosynthesis by repressing the expression of *fabA* and *fabB*. As low levels of ω7-UFAs are required for cell growth, we reduced *fabA* and *fabB* expression using CRISPR interference (CRISPRi)(37, 48).

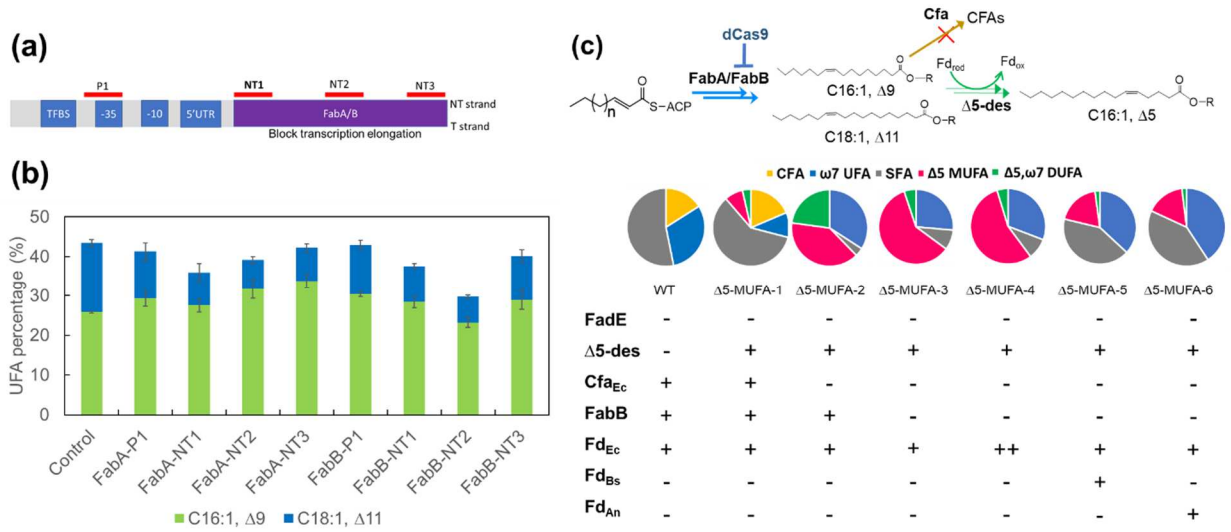


Figure 3. 5. Engineering E. coli to produce Δ5-MUFA in E. coli. a. sgRNAs bind to different regions of non-template strand of fabA or fabB gene. b. CRISPRi with different sgRNA showed different inhibition efficiency for the production of ω7-UFAs. Targeting of sgRNA to the middle region of fabB gene obviously inhibit the production of ω7-UFAs. Error bars represent standard deviation measured from biological triplicates. c. The production of Δ5-MUFA in different engineered strains.

Small guide RNAs (sgRNAs) targeting four different regions of both the *fabA* and *fabB* operons were designed to search for the most optimal level of repression (Figure 3.3a). Constitutively expressed sgRNAs were introduced on a plasmid to ω7-UFA strains harboring a *dCas9* gene. Fermentation of the resulting strains showed that a sgRNA targeting nucleotides 484-504 (approximately the middle) of the *fabB* coding region was the most effective in decreasing ω7-UFA composition (Figure 3.3b). This sgRNA was then expressed in the Δ5-MUFA strain. The resulting strain (Δ5-MUFA-3) has 60.0±0.1% of Δ5-MUFA in phospholipid FAs, with only 26.4±0.4% native ω7-UFA. Additional overexpression of ferredoxin from either *E. coli*, *B. subtilis*, and *Anabaena* sp. 7120 reduced Δ5-MUFA composition (Figure 3.3c, Figure 3.9S), suggesting an imbalanced redox potential.

3.2.2 Phospholipid profile analysis of the IBFA-producing strain

After obtaining diverse FA profiles in phospholipids, it is interesting to know how these

uncommon FAs are distributed between different phospholipid species, as that can illuminate the substrate specificity of FA-modifying enzymes. In wild-type *E. coli*, phospholipid biosynthesis starts from acylation of glycerol-3-phosphate at both sn-1 and sn-2 positions to form phosphatidic acid (PA). PA is then converted to CDP-diacylglycerol (CDP-DAG) followed by exchanging the head group to phosphatidylethanolamine (PE, ~70%), phosphatidylglycerol (PG, ~25%), cardiolipin (CL, ~4%), and phosphatidylserine (PS, 0.1% of phospholipid)(49) (Figure 3.4a)(29, 50). While the phospholipid substrates for CFA and Δ5-desaturase have been elucidated in previous studies(5, 51, 52), phospholipid substrates for the IBFA pathway remain unknown. To identify the profiles of IBFA on different phospholipids and their sn-positions, we performed lipid analysis of the IBFA-rich strain (IBFA-5) using liquid chromatography-mass spectrometry (LC-MS/MS). The results showed that there was no detectable IBFA in PA, CDP-DAG, and CL. The percentage of IBFA in PG, was significantly higher than PE, or PS: 57.0%±0.17, 41.0%±1.01, and 39.1%±2.63 (adjusted p values < 0.05) respectively (Figure 3.4b).

Further analyses of the sn-positions of IBFAs identified a higher methylation activity at the sn-1 position than that of the sn-2 position for each phospholipid (Figure 3.4c). At the sn-1 position, 95.4%±0.2 of ω7-UFAs were converted to IBFAs in PG, and 85.0%±2.8 and 79.0%±0.7 of conversion were obtained for PE and PS, respectively.

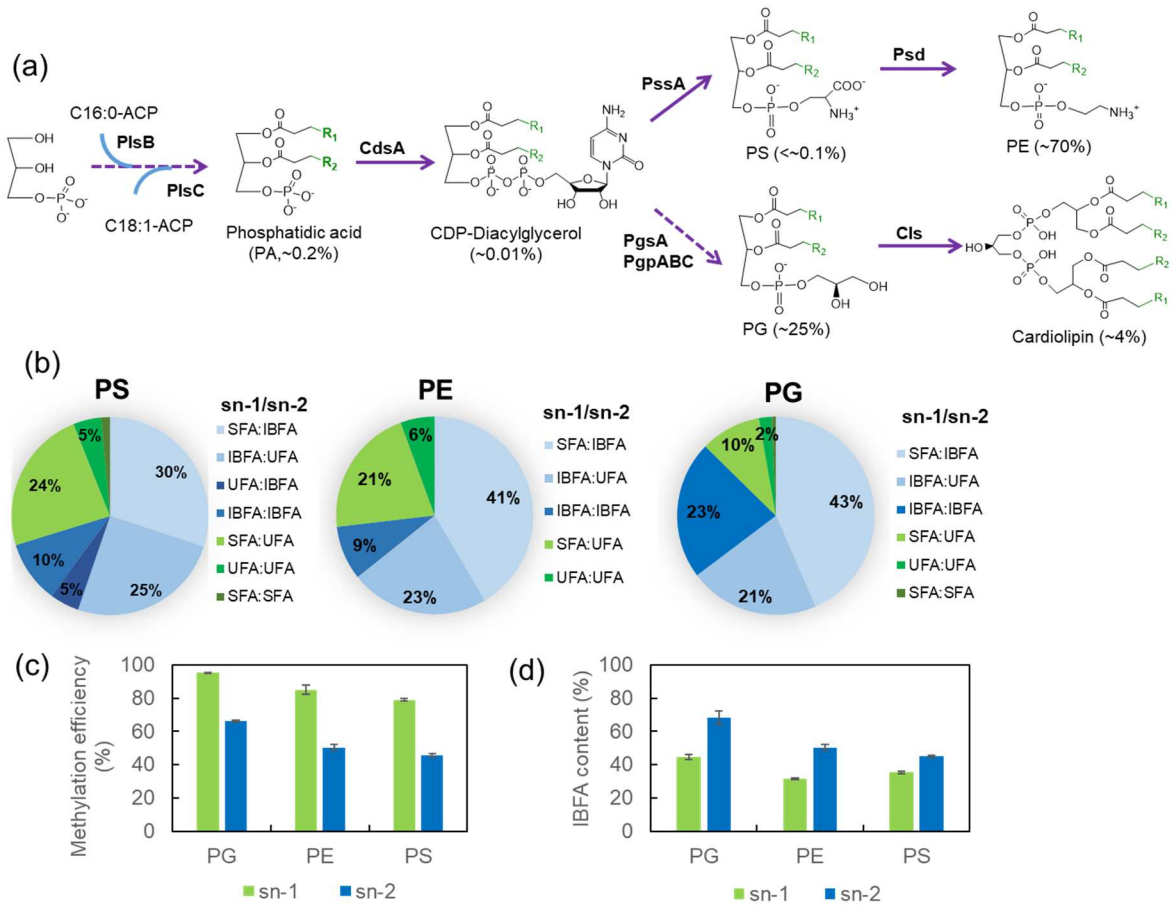


Figure 3. 6.Lipid profile analysis of the IBFA-enriched strain IBFA-5. a. The biosynthesis pathway of major phospholipid species in E. coli. b. The distribution of IBFA in different phospholipid species in engineered strain. c. The methylation efficiency of BfaB as calculated from the ratio of IBFAs to UFA equivalent in different phospholipid species. d. The content of IBFAs in different position of different phospholipids. PS: Phosphatidylserine; PE: Phosphatidylethanolamine; Phosphatidylglycerol: Phosphatidylglycerol; Cds: phosphatidate cytidylyltransferase; Pss: phosphatidylserine synthase; Psd: phosphatidylserine decarboxylase; PgsA: phosphatidylglycerolphosphate synthase; PgpA/B/C: phosphatidylglycerolphosphate phosphatase; Cls: cardiolipin synthase; CDP-DAG: cytosine diphosphate-diacylglycerol.

At the sn-2 position, the ω7-UFA to IBFA conversion ranged from 45.7%±1.1, 50.4%±1.65, 66.5%±0.5 for PE, PG, and PS respectively (Figure 3.4c), suggesting the IBFA pathway prefers ω7-UFAs at the sn-1 position of phospholipids. However, the total contents of IBFAs at the sn-1 position (PE = 31.6%±0.5, PG = 44.5%±1.5, and PS = 35.4%±0.9) are lower than that at the sn-2 position (PE = 50.4%±1.6, PG = 68.1%±4.1, and PS = 45.0%±0.7) for each phospholipid (adjusted

p values < 0.05) (Figure 3.4d).

3.2.3 Phenotypic profiling of engineered strains with diverse FA profiles

Engineered bioproduction strains often suffer from reduced cell growth and altered metabolism in comparison to their wild-type counterparts(53-55). Here we have created several strains with drastically different FA profiles in phospholipids compared to that of wild-type *E. coli*. We next examined whether these modified FA profiles would affect metabolic activity and cell growth. To test this, we performed phenotype microarrays by cultivating the engineered strains which produced the highest titers of ω7-UFA (ω7-UFA-5), CFA (CFA-4), IBFA (IBFA-5), and DUFA (DUFA-5) in 96 different growth media and conditions such as different carbon sources, organic acids, salts, pH, reducing power, and antibiotics (Figure 3.11S). Screening of these strains revealed that the IBFA- and CFA-enriched strains have similar growth and metabolic characteristics compared to the control *E. coli* strain without phospholipid FA modification (Figure 3.5a, Table 3.5S, Table 3.6S).

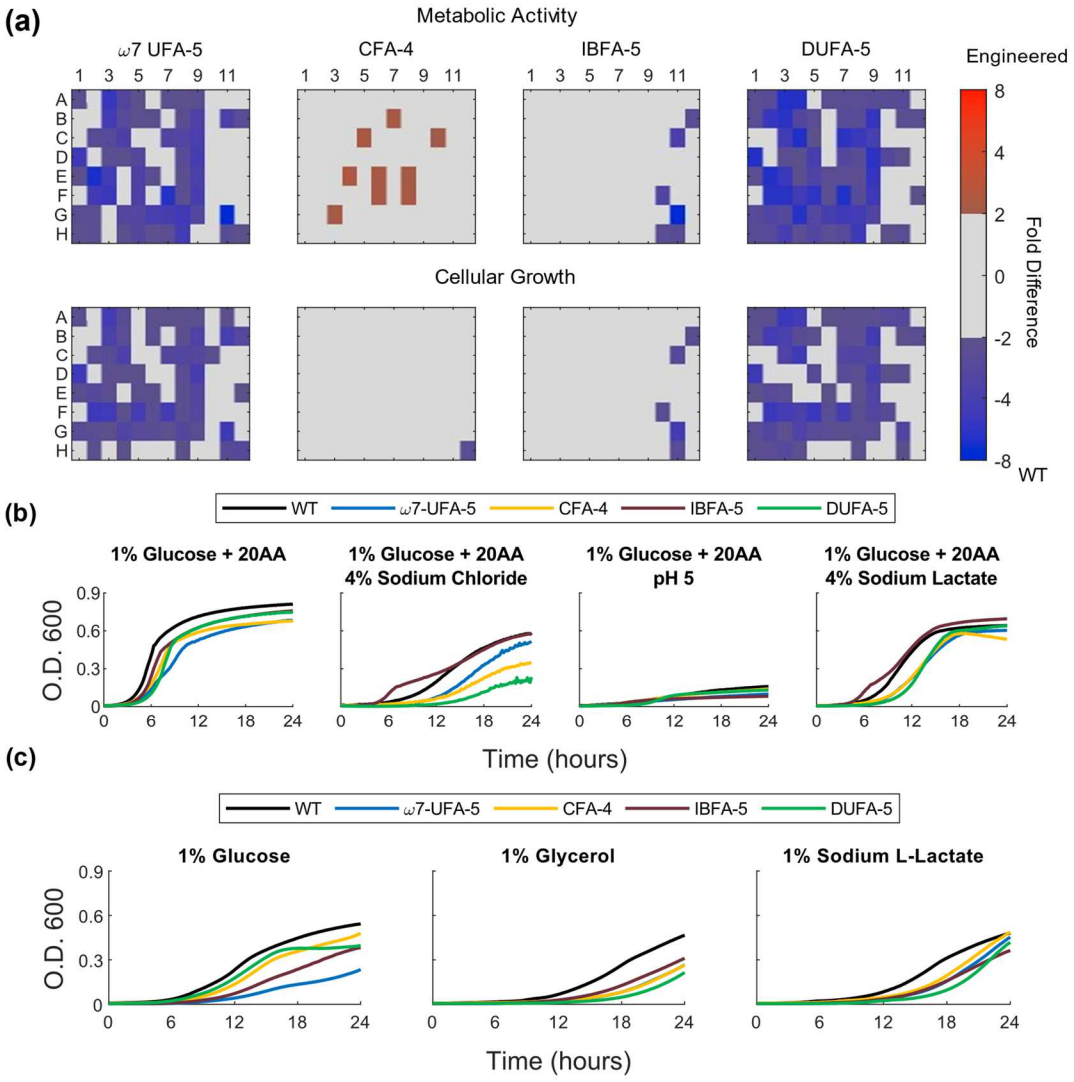


Figure 3. 7 Phenotypic profiling of engineered strains with diverse FA profiles. a. Metabolic activity (top) and Cellular Growth (bottom) across 96 metabolic conditions. Measured Activity is compared against WT strain measured under same conditions. b&c. Average growth curves of biological triplicates in select challenge conditions of engineered strains (WT, black; ω7-UFA-5, blue; CFA-4, yellow, IBFA-5, purple; DUFA-5, green).

Out of the 96 tested growth media and conditions, the IBFA- and CFA-enriched strains have more than 2-fold difference in cell densities in only 5 (reduced growth in 8% NaCl, fusidic acid, vancomycin, lithium chloride, and sodium butyrate) and 1 (sodium bromate) conditions, respectively (Table 3.6S). Metabolic activities under these conditions were measured using a tetrazolium redox dye to track oxidative phosphorylation activity. Both IBFA- and CFA-enriched

strains exhibited similar metabolic activity to those of the control *E. coli* strain, with only 6.3% and 10.4% of conditions displaying more than 2-fold difference. Interestingly, only the CFA-enriched strain exhibited any increased metabolic activity, under 10 conditions: Dextrin, N-acetyl-β-D-Mannosamine, 3-Methyl Glucose, 1% sodium lactate, L-arginine, L-Glutamic acid, L-Pyroglutamic Acid, Glucuronamide, Quinic acid, D-Lactic acid Methyl Ester (see Table 3.5S for all conditions). Of the 10 conditions with enhanced respiratory activity for CFA, 6 are carbon source conditions which can also be used as a nitrogen source (N-acetyl-B-D-Mannosamine, L-arginine, L-glutamic acid, l-pyroglutamic acid, Glucuronamide), or are amino acid precursors (Quinic acid). This did not correlate with any increased growth in those conditions. The ω7-UFA- and DUFA-enriched strains had reduced metabolic activities and lower cell density for 59.4% (57/96) and 74.0% (71/96) of the tested conditions, respectively.

Additionally, these strains were cultivated in parallel under different stress conditions (Figure 3.5b, Figure 3.12S) and in different commonly used carbon sources (Figure 3.5c, Figure 3.12S). Mutant strains reached an estimated maximal growth rate lower than WT when grown in glucose, glucose supplemented with amino acids, and high salt. Mutant strains reached similar maximal growth rates compared to the control strain when grown in glucose supplemented with sodium lactate, sodium L-lactate, or in low pH, or grown in glycerol (Figure 3.13S, Figure 3.14S), confirming that these modifications in phospholipid FAs are tolerated by the *E. coli* chassis under controlled laboratory conditions. There were no significant differences between the maximal cell density reached by WT and any of the mutant strains (Figure 3.15S, Table 3.7S).

3.2.4 Transcriptomic profiling of engineered strains with diverse FA profiles

Given the aforementioned difference in metabolic burden imposed on the *E. coli* chassis between the IBFA-, CFA-, and DUFA-enriched strains, we next investigated genome-wide

expression changes in all strains compared to the UFA-enriched strain (control). To identify the effect of altering FA composition on the transcriptome at different growth phases, we sampled liquid cultures of each strain at 3 time points: 4, 8, and 24 hours. A biplot of the principle component analysis of the 500 most variable genes shows that ellipses generated from control, IBFA, and CFA conditions vary along principle component PC2 (Figure 3.16S). This confirms that the transcriptional state of the UFA, CFA, and DUFA expressing strains were very similar across time (adjusted p value > 0.05 [PERMANOVA]). The DUFA samples did not follow this relationship, instead varying along PC1, with the DUFA-enriched samples at 4- and 8- hours clustering together away from samples of other strain at similar time points. PERMANOVA comparison of all DUFA samples compared to samples of other strains confirmed a significant difference in variance over time (adjusted p value < 0.05 versus all other strains). Additionally, examining the 40 most variable genes across time identified marked upregulation of the *valU*, *valX*, *valY*, *lysZ*, *lysQ*, and *lysY* genes, all members of the aminoacyl-tRNA synthesis pathway, (Figure 3.17S) in the DUFA-enriched strain.

Strain 1	Strain 2	R2	pval	pvalBon	pvalFDR
32	34	0.100251	0.154	0.924	0.185
32	35	0.076084	0.23	1.38	0.23
32	36	0.240502	0.007	0.042	0.014
34	35	0.11853	0.106	0.636	0.159
34	36	0.314532	0.001	0.006	0.006
35	36	0.238559	0.004	0.024	0.012

Table 3. 1 Permanova tests of biplot PCA Permuted anova of transcriptomic samples from each FA producing strain. Analysis of variance for all strains was conducted, and then either Bonferroni or FDR hypothesis testing correction was implemented on each p value to correct for false positives. A corrected p value under 0.05 indicates the sampled transcriptomes of the two tested strains are significantly different.

In order to understand the effects of these alterations on the *E. coli* chassis, we compared

the most highly significant and differentially abundant transcripts within the *E. coli* genome (here defined as < 0.01 adjusted p [Deseq2] and > 3 log fold-change) for the CFA, IBFA, and DUFA strains versus UFA, and compared the shared and unique genes within these sets at 4 hours, during mid-log growth phase (Figure 3.6A).

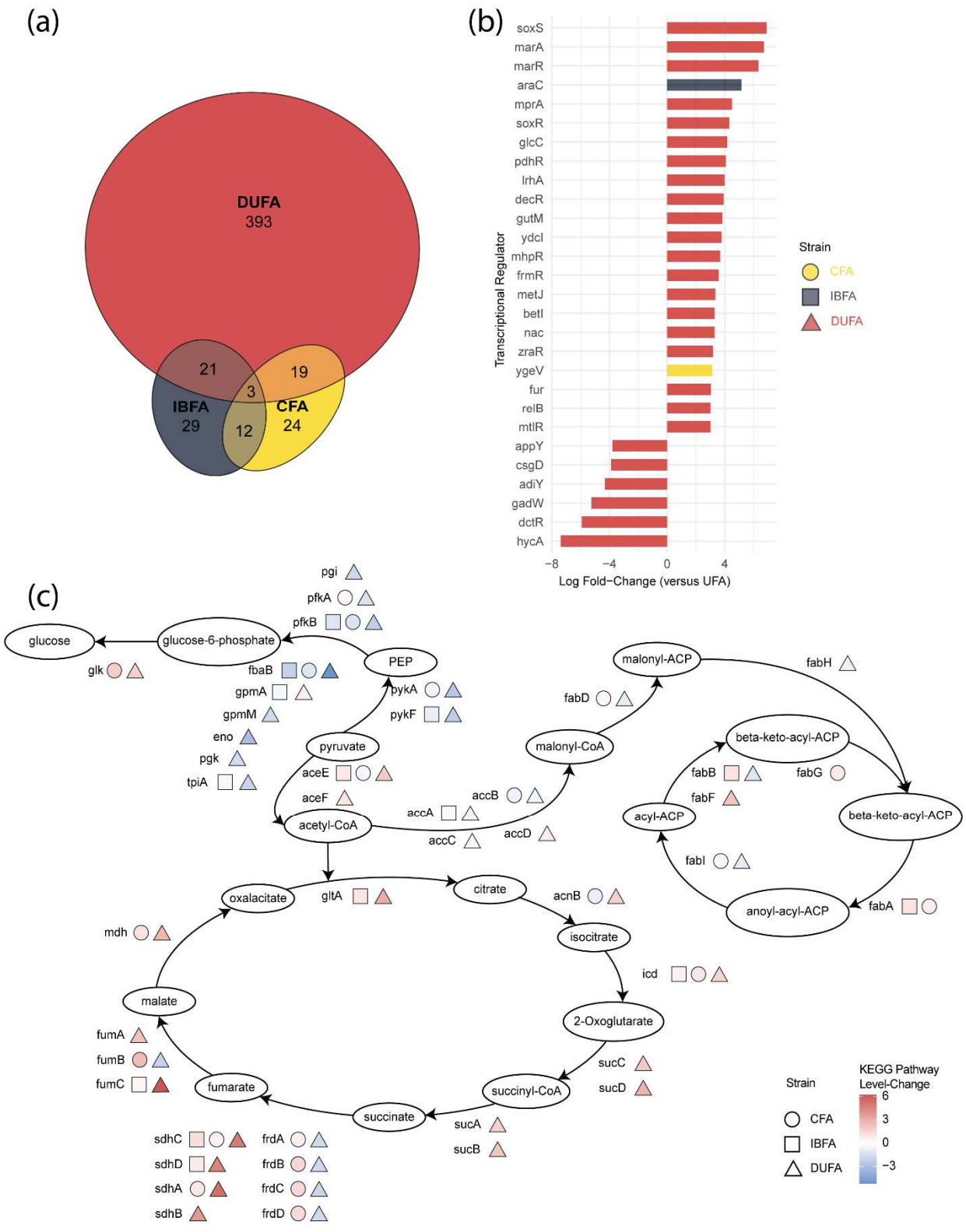


Figure 3. 8. Transcriptomic profiling of engineered strains with diverse FA profiles. a. A Euler diagram showing the intersect and union of the highly DE genes in CFA, IBFA, and DUFA producing strains compared to UFA at 4 hours. b. Barplot showing the log fold change for every transcriptional regulator

within the set of highly DE genes unique to one strain from 6. Color represents the strain they were found to be highly DE in. c. Map of the fatty acid biosynthesis, TCA cycle, and glycolysis pathways. The shapes next to each gene of the pathway represent the three strains, while the color represents the magnitude and direction of differential expression.

Over 78% (393) of the high differentially expressed (DE) genes were unique to the DUFA, while there were only 3 genes shared amongst all strains: *ybdL, metF, and metA*. Comparing the genes unique to each comparison, we found that CFA and IBFA each had less than 30 unique DE transcripts compared to UFA, among which were only two transcriptional regulators: *ygeV* (CFA) and *araC* (IBFA) (Figure 3.6B). Comparatively, DUFA overexpression resulted in 26 DE transcriptional regulators. There was significant upregulation of genes known to produce stress-related transcriptional cascades such as *soxR/S, marA/R, betL,* and *zraR.* We also observed DE in several transcriptional regulators known to affect metal homeostasis, such as overexpression of *fur*, the ferric uptake regulator, and *metJ*, which drives repression of the sulfurous amino acid methionine. Further investigation revealed that the iron acquisition/uptake genes regulated by *fur* were differentially expressed (Figure 3.18S). Finally, we examined the DE in genes from the central metabolic pathways of glycolysis, citrate (TCA) cycle, and fatty acid biosynthesis for differences between the UFA and the other strains at 4 hours (Figure 3.6C). As was seen with the differential expression analysis, the DUFA-enriched strain exhibited strong DE in these pathways (37 of 89 total genes, or 42%), specifically in TCA cycle genes, including upregulation of the complete *sdh* operon (also a part of the *fur* regulon, see Figure 3.18S), and downregulation of many genes in the glycolysis pathway. IBFA and CFA production only resulted in 17 (19%) and 23 (26%) of genes within the three pathways, respectively.

3.2.5 Engineering *E. coli* for production of free IBFAs

To demonstrate these engineered strains are useful for production of non-common FAs, we attempted to engineer the high IBFA content strain (IBFA-5) to produce free IBFAs (FIBFAs).

FIBFAs can be readily converted to esters(7), alkanes(21), or alcohols(56) with drastically reduced melting temperature compared to their straight-chain or even terminally-branched counterparts and used as jet fuels or low-temperature lubricants(6). Production of FIBFA in engineered microbial hosts have not been previously reported.

FIBFAs can be released from phospholipid by phospholipases that hydrolyze the ester bonds in glycerophospholipids(57). Wild-type *E. coli* has one outer-membrane-bound phospholipase PldA (encoded by *pldA*) and an inner-membrane-bound phospholipase PldB (encoded by *pldB*, Figure 3.7a)(58).

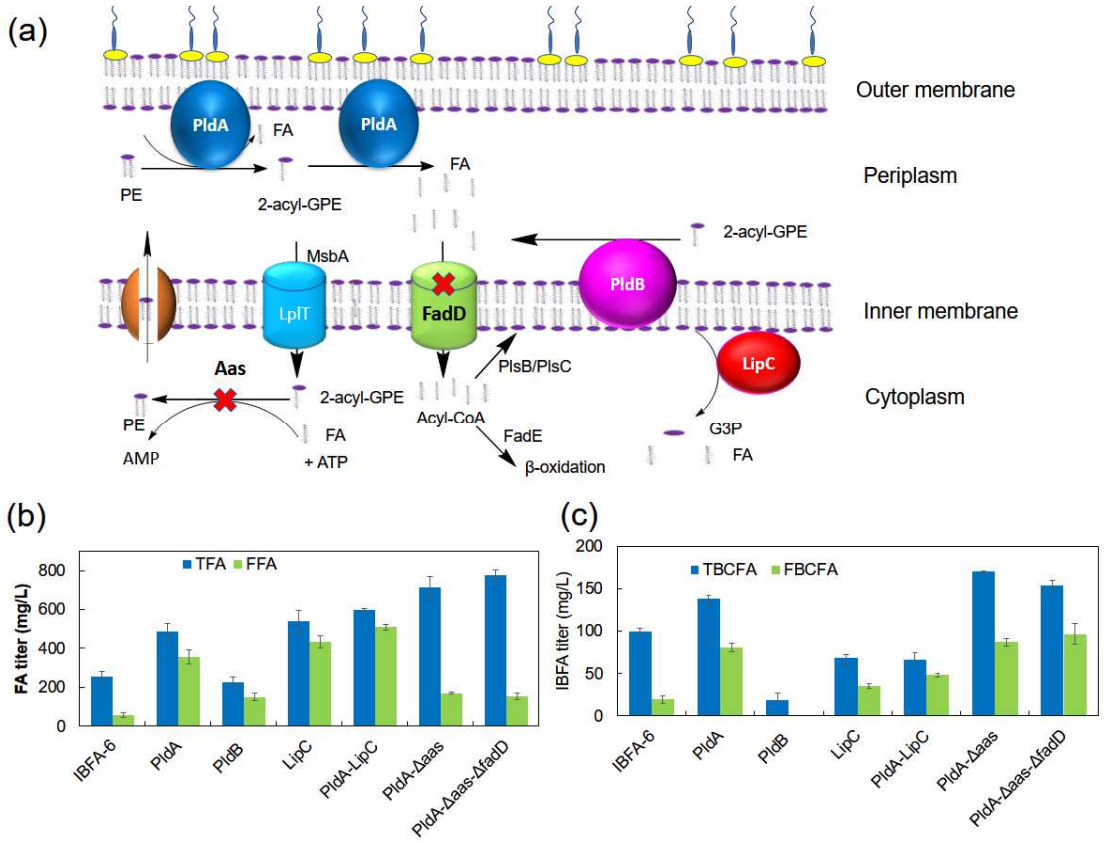


Figure 3. 9.FIBFA production in engineered E. coli. a. Metabolic engineering strategies for the production of FIBFAs from phospholipids in E. coli. b. The titer of total FAs (TFAs) and FFAs in engineered strains. c. The titer of total IBFAs (TIBFAs) and FIBFAs in engineered strains. Error bars represent standard deviation measured from biological triplicates.

PldA and PldB specifically hydrolyze the FA ester bond at either sn-1 or sn-2 position of phospholipids, respectively(57). Additionally, the inner-membrane-associated phospholipase LipC (encoded by *lipC*) from *B. subtilis* can hydrolyze the FA ester bond at both sn-1 and sn-2 positions(59, 60). We first separately overexpressed *pldA*, *pldB*, and *lipC* in the IBFA-enriched strain IBFA-5 that contains 40% IBFAs in phospholipids. The fermentation results showed that expression of *pldA*, *pldB*, or *lipC* produced 355.3 mg/L, 150.6 mg/L, and 433.3 mg/L of free FAs (FFAs), respectively, confirming their phospholipase activities (Figure 3.7b). While the *pldA*-expressing strain has a slightly higher IBFA fraction in phospholipids (44%, v.s. 40% in the strain without phospholipase overexpression), the *pldB*- and *lipC*-overexpressing strains have lower IBFA fraction in phospholipids (24% and 33%, respectively). This is likely caused by the spatial competition between the methyltransferase BfaB with *pldB* or *lipC* for access to inner membrane phospholipids(58, 59). As inner-membrane-bound or -associated proteins, *pldB* or *lipC* may block BfaB from methylating phospholipids in the cytosolic compartment. When free IBFAs were quantified, while the *pldB*-overexpressing strain failed to produce any FIBFA, the *pldA* and *lipC*-overexpressing strains produced 80 mg/L and 35 mg/L FIBFAs, respectively. Next, *pldA* and *lipC* were coexpressed. Although the titer of total FFAs was increased to 509 mg/L, FIBFA titer was decreased compared to the strain only expressing *pldA*, further suggesting that overexpression of inner-membrane-bound proteins hinder BfaB from methylating phospholipids. Thus, we focus on the strain that only expressed *pldA*.

Next, we targeted to the transmembrane allocation of FFAs. As an outer-membrane enzyme which faces the periplasm, PldA hydrolyzes PE to 2-acylglycerophosphoethanolamine (2-acyl-GPE) and FFAs, both of which could be transported into the cytosol. In the cytosol, acyl-ACP synthetase (AAS, encoded by *aas*), has 2-acylglycerophosphoethanolamine acyltransferase

activity and can acylate 2-acyl-GPE back to PE(61, 62) (Figure 3.7a). To prevent this futile cycle, the *aas* gene was deleted, resulting in increased titers for both FFA (by 29% to 507.0 mg/L) and FIBFAs (by 12% to 86.9 mg/L) (Figure 3.7d). Additionally, periplasmic FFAs can be transferred into the cytosol and simultaneously activated to acyl-CoA by FadD (encoded by *fadD*). Although acyl-CoA cannot be degraded in our engineered strain due to the deletion of the β-oxidation enzyme FadE, it can be incorporated back into the phospholipids in the inner membrane by PlsB and PlsC(50, 61) (Figure 3.7a). To prevent this, *fadD* was also deleted. The resulting strain FIBFA-8 further increased FFAs up to 623.4 mg/L and FIBFAs up to 96.3 mg/L (Figure 3.7c).

3.3 Materials and Methods

3.3.1 Reagents

Phusion DNA polymerase was acquired from New England Biolabs (Beverly, MA, USA). Restriction enzymes and T4 DNA ligase were purchased from Thermo Fisher Scientific (Waltham, Massachusetts, USA). Oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, IA, USA). Standard 10-methyloctadecanoic acid and 10-methylhexadecanoic acid were purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA) and Matreya company, (State college, PA, USA), respectively. Bacterial acid methyl ester mixture standards, C4-C24 even carbon saturated FAMEs, GLC-50, GLC-90, and all the other reagents were purchased from Sigma Aldrich (St. Louis, MO, USA).

3.3.2 Strains and plasmids

All strains and plasmids used in this study are listed in Table 3.2S and 3.3S. *E. coli* DH10β was used for gene cloning and *E. coli* MG1655Δ*fadE* was used for FA production. *E. coli* and *B. subtilis* genes were amplified from their genome by colony PCR using primers listed in Table 3.4S. All

other genes were codon-optimized for *E. coli* expression and chemically synthesized by Integrated DNA Technologies (Coralville, IA, USA). All plasmids were constructed using standard Golden-Gate DNA assembly method(63). The *cfa*, *fabR, aas* and *fadD* genes were deleted from *E. coli* genome following a CRISPR/Cas9 gene replacement method(64). All deletions were confirmed by colony PCR. Repression of *fabA* and *fabB* genes was achieved using a CRISPR interference (CRISPRi) method, by expressing constitutively sgRNAs from plasmids(48). Plasmids were transformed into electro-competent strains by electroporation and selected on LB agar plates with proper antibiotics (ampicillin, 100 mg/L; kanamycin, 50 mg/L; chloramphenicol, 34 mg/L; spectinomycin 60 mg/L).

3.3.3 Fermentation

For each transformant, 3 different colonies were selected and used to inoculate 3 ml LB medium with proper antibiotics. The overnight cultures 1% (v/v) were then used to inoculate 20 ml of modified M9 minimal medium (containing 75 mM MOPS at pH7.4, 2 mM $MgSO_4$, 1 mg/Lthiamine,10 μM $FeSO_4$, 0.1 mM $CaCl_2$ and micro nutrients including 3 μM $(NH_4)_6Mo_7O_{24}$, 0.4 mM boric acid, 30 μM $CoCl_2$, 15 μM $CuSO_4$, 80 μM $MnCl_2$, and 10 μM $ZnSO_4$) with 2% glucose, 0.5% yeast extract, and corresponding antibiotics in 250 ml flask. When OD_{600} reached 0.6-0.8 at 37 ºC, inducer (0.8% Arabinose, 0.5 mM Isopropyl β-D-1-thiogalactopyranoside(IPTG), 2 μM anhydrotetracycline) was added and the cell cultures were grown at 18°C for 48 h (ω7-UFA, CFA, DUFA) or 30°C for 24 h (Δ5-MUFA, IBFA). After fermentation, cells were harvested for FA quantification.

3.3.4 Quantification of total FAs

The amount of total FAs was quantified using previously published method(65). In detail, 1 ml of cell culture was pelleted and added with 1 ml chloroform, 1 ml of 15% (v/v) H_2SO_4/methanol, and

40 mg/L of nonadecanoic acid as an internal standard and the mixture was heated at 100°C for an hour for transesterification. Reaction mixture was then cooled on ice for 5 min, followed by the addition of 1ml purified water followed by vigorous shaking for 1 min (30 sec/once, twice). The organic phase containing fatty acid methyl esters (FAMEs) was isolated and directly analyzed using a gas chromatograph (GC) (Hewlett-Packard model 7890A, Agilent Technologies) equipped with a 30 m DB5-MS column (J&W Scientific) and a mass spectrometer (5975C, Agilent Technologies) or a Flame Ionized Detector (FID) (Agilent Technologies). The column was equilibrated at 80°C for 1 min, followed by a ramp to 300°C at 20 °C/min, and was then held at 300°C for 3 min. FA species were identified by comparing their retention times to those of standard FA methyl esters (Bacterial Acid Methyl Ester Mix, Sigma Aldrich) and by comparing their mass spectra to the Probability Based Matching (PBM) Mass Spectrometry Library. C18:2, Δ5Δ11 was identified by comparing GC-MS spectra of efedrenic acid methyl ester to the published spectra(47). FA concentrations were quantified by comparing the area of each FAME peak to a standard curve generated using standard FAME mixtures (C4-C24 even carbon saturated FAMEs, GLC-50, GLC-90). The titer of FAs for each strain was measured in biological triplicates.

3.3.5 Quantification of membrane lipids

Membrane lipids were extracted using Bligh and Dyer lipid extraction protocol(66). In detail, cells from 3 ml culture were collected and washed with 0.6% LiCl for 3 times. The pellet was resuspended into 3.8 ml of a MeOH/ H_2O /$CHCl_3$ (2:0.8:1, v/v/v) solution, followed by sonication for 30 sec on ice. Then, 1 ml $CHCl_3$ and 1 ml 0.63% LiCl were added into the cell lysate. After vigorous shaking for 30 sec, the mixture was centrifuged at 800 g for 5 mins. The bottom chloroform phase that contains the extracted lipids was transferred into a GC vial using a long Pasteur pipet. The remaining aqueous solution was extracted again using 1 ml of fresh chloroform.

To determine the amount of FAs in membrane lipids, the extracted FAs were converted to FAMEs using the above-mentioned method, followed by GC-FID quantification.

3.3.6 High resolution mass spectrometric analysis of IBFA strain cell membrane

Profiling and structural characterization of phospholipids with high resolution (R = 100,000 at m/z 400) mass spectrometry were performed on a Thermo Scientific (San Jose, CA) LTQ Orbitrap Velos mass spectrometer (MS) with a built-in syringe pump system operated by Xcalibur operating system. A 50ul/min solvent (methanol with 0.5 % NH_4OH) was continuously infused into the ESI source and samples in methanol (10 ul) were flow injected. The skimmer of the ion source was set at ground potential, the electrospray needle was set at 4.0 kV, and temperature of the heated capillary was 300°C. The automatic gain control of the ion trap was set to $5x10^4$, with a maximum injection time of 100 ms. Helium was used as the buffer and collision gas at a pressure of $1x10^{-3}$ mbar (0.75 mTorr). The MS^n experiments were carried out with an optimized relative collision energy ranging from 25-40% and with an activation q value at 0.25. The activation time was set for 10 ms to leave a minimal residual abundance of precursor ion (around 20%). The mass selection window for the precursor ions was set at 1 Da wide to admit the monoisotopic peak to the ion-trap for collision-induced dissociation (CID) for unit resolution detection in the ion-trap or high-resolution accurate mass detection in the Orbitrap mass analyzer. Mass spectra were accumulated in the profile mode, typically for 3-10 min for MS^n spectra (n=2,3,4). The structural assignments of phospholipid species are based on the MS^n spectra previously described(67)

3.3.6 Quantification of FFAs

The concentration of FFAs was determined using the previous published method(26). In detail, 0.5 ml of cell culture was harvested and acidified with 50 µl of 12 N HCl., The mixture was extracted twice with 0.5 ml ethyl acetate, which was spiked with 20 µg/ml of nonadecanoic acid (C19:0) as

an internal standard. FFAs were derivatized to FAMEs by adding 90 µl methanol, 10 µl 12 N HCl, and 120 µl trimethylsilane-diazomethane. The mixture was incubated at room temperature for 10 min. FAMEs were analyzed using GC-FID.

3.3.7 Metabolic Profiling Assays

Colonies from fresh transformation were used to inoculate 3 ml of LB medium with relevant antibiotics and grown at 37ºC overnight. The overnight culture was used to inoculate 5 ml of modified M9 minimal medium supplemented with 2% glucose, 0.5% yeast extract, 0.8% arabinose (inducer) and corresponding antibiotic. Cells were incubated at 18°C for 42 hours (ω7-UFA, CFA, DUFA), or 30°C for 18h (WT, IBFA). Cells were diluted 1% in fresh modified M9 and fermentation continued for another 6 hours to mid-exponential growth (OD_{600} at 0.6-0.8). For metabolic profiling assays, cells were harvested and washed twice in PBS via centrifugation at 4000 rcf, 2 minutes, then resuspended in Inoculating Fluid-A (IF-A, Biolog, Hayward, California, USA) with 0.008% arabinose and corresponding antibiotic. Cells were transferred to a GEN III microplate (Biolog, Hayward, California, USA) and cultivated from a starting OD_{600} of 0.01 at 30°C (for all strains) for 16 hours. Absorbance measurements at 590 nm (metabolic activity) and 750 nm (cell growth) were made using an Infinite F200PRO plate reader (TECAN, Männedorf, Switzerland).

3.3.8 Time course growth assays

For time course growth assays, cells were harvested and washed twice in modified M9 with 0.04 arabinose and 5x antibiotic without any other carbon source. Cells density was normalized to OD_{600} of 0.2, then diluted 5-fold into modified M9 supplemented with carbon, amino acids, and/or harsh condition challenges in a 96-well imaging microplate (Corning, Corning, New York, USA). For cultures with amino acids, the concentrations of the 20 amino acids were standard for EZ rich medium. Plates were measured and incubated in an Infinite F200PRO at 30°C with 3 mm orbital

shaking. Absorbance measurement at 600 nm were made every 10 minutes for 24 hours. Data was imported into R (v4.1.0). Max OD was calculated for each replicate using the max() function in R and averaged across triplicates for each condition. Maximal growth rate was estimated by taking the first derivative of the growth curve and fitting a loess regression. The span which resulted in the lowest sum of squared error was selected for each regression, and the max value of that regression and the confidence interval calculated at that point was used for comparisons between strains.

3.3.9 Transcriptomics analysis

Sample Collection

The high IBFA content strain IBFA-5 and the control strain ω7-UFA-5 were cultivated in modified M9 minimal medium as described above. After induction for 4 h, 8 h, and 24 h, 2 ml of cell culture was harvested. Cell pellet was collected by centrifuge at 13,000 g for 2 min, and gently resuspended in 500 µl RNAlater buffer. The resuspension was incubated at room temperature for 20 minutes and immediately frozen in liquid nitrogen.

RNA extraction and cDNA synthesis

Total genomic material (DNA/RNA) was extracted from frozen cell suspensions using the Quick-RNA fungal/Bacterial Miniprep Kit (Catalog #R2014) from Zymo Research. Cell lysis was conducted using a MiniBeadBeater 24 manufactured by Biospec products for 1.5 minutes at medium speed. DNA was removed using a modified protocol from the TURBO DNase kit: one round of digestion was immediately followed by spiking in the same amount of DNase and another successive round of heat treatment. Samples were then concentrated and purified using the RNA Clean & Concentrator (Catalog #11-325). Sample total RNA concentrations were measured using a NanoVue (General Electric). We checked for gDNA contamination via PCR of 0.5µl of sample using a primer sequence designed to target 1640 base pair section of *E. coli gyrA* (Forward:

tcttccaggttgatgtctgc, Reverse: tttgcgacctttgaatccgg) samples with pcr products underwent another round of TURBO DNase treatment. We used the Illumina Ribo-Zero Magnetic Kit to deplete rRNA, after which mRNA samples were cleaned and purified using the Agencourt RNAClean XP. First Strand cDNA synthesis was performed using SuperScriptII reverse transcriptase, Second strand cDNA synthesis used *E. coli* DNA polymerase I (New England Biolabs) and RNase H (New England Biolabs) and used *E. coli* ligase (New England Biolabs) and RNase H. Final mRNA concentrations were quantified using the Qubit DNA HS protocol on a Qubit 4 Fluorometer (ThermoFisher). cDNA was stored at -20°C until sequencing library creation.

cDNA sequencing

Shotgun metagenomic sequencing libraries were created using cDNA diluted to 0.5 ng/µl and the modifications to the Nextera library prep lot (Illumina) detailed in Baym et al(68). This generated ~450 bp DNA fragments which were purified using the Agencourt AMPure XP system (Beckman Coulter) and quantified using the Quant-it PicoGreen dsDNA assay (Invitrogen). Samples were pooled onto lanes to ensure ~10M (1x75bp) reads per sample and quantified using the Qubit dsDNA BR assay. Samples were sequenced in Illumina NextSeq High-output sequencing machines at the Edison Family Center for Genome Sciences and System Biology at Washington University School of Medicine in St. Louis.

Data Analysis

The pipeline for transcriptomic analysis proceeded similarly as previously described(69). Briefly, samples were trimmed using trimmomatic and then mapped to a bowtie2(70) index built using the *E. coli* MG1655 NCBI assembly GCF_000005845.2_ASM584v2 including a custom plasmidic operon containing the novel biosynthetic pathways for each strain. Gene counts were calculated using featureCounts(71). Differential expression analysis was performed using DESeq2(72), comparing the UFA-enriched strain as the comparator group for estimating differential expression

in the CFA- IBFA- and DUFA-enriched strains. Repeated PERMANOVA conducted in R using the Adonis function of vegan(v2.5-7)(73) package comparing all timepoints of all strains to each other (999 permutations, Euclidean distance) after rlog normalization in DESeq2. Pathway analysis was conducted using the Escherichia coli K-12 MG1655 KEGG pathway gene lists, and testing for significant log fold-change in expression among triplicates using DESeq2.

3.4 Discussion

In altering the FA-profiles of *E. coli*, we produced significant increases in multiple forms of highly desirable precursor molecules. In addition, we gained substantial knowledge pertaining to the effects of altered FA composition on metabolism, growth, and the transcriptional network of the cell. Increasing ω7-UFA biosynthesis is a necessary precursor to synthesis of CFAs and IBFAs, and thus a similar strategy was applied to all the strains produced in this study: 1) isolating FA biosynthesis from the degradation pathway via deletion of *fadE,* 2) deletion of the native CFA biosynthesis gene, 3) deletion of *fabR*, and 4) overexpression of *fadR*, the main transcriptional regulator for FA biosynthesis(40). Surprisingly, overexpression of *fabA* or *fabAB* yielded similar relative proportions of ω7-UFA. The lack of difference between the two suggest that the rate-limiting step of ω7-UFA biosynthesis is catalyzed by FabB, rather than FabA, which is consistent with previous reports(40). Overexpression of *fabB* or *fabR* has been attempted before(74), and in one study produced a similar percentage of total UFA content(4); however, it is unclear whether a similar yield of UFA is possible without +*fadB*/Δ*fabR* in tandem. In the end, deletion of *fabR* alone produced the highest ω7-UFA titers of all engineered strains (Figure 3.2a),(40) suggesting that regulatory control of *fabR* is extremely tight and can continue to limit overall FA production even when FA synthesis has been overexpressed. A comparison of our engineering strategies with previous works for increasing UFA fraction is provided in Table 3.7S.

Once the MG1655Δ*fadE*Δ*cfa*Δ*fabR*+*fadR* genetic chassis was established, strains optimized for increased CFA, IBFA, and DUFA synthesis were produced by introduction of only two extra genes each. Though cyclopropanation increased the metabolic activity of *E. coli* in multiple tested carbon sources, increased CFA production did not result in improved growth rate when grown in high salt conditions, suggesting overproduction of CFA still imposes a significant burden on the cell when compared to WT; instead, it performed similarly to the other engineered strains. Interestingly, the CFA overproducing strain displayed fewer growth defects compared to strains with increased UFAs and DUFAs (Figure 3.5a), yet still displayed a similar expression profile to the UFA strain at all time points (Figure 3.17S). These results indicate that the expression profile shared by increased UFA, CFA, and IBFA synthesis could be driven by increased energy demand versus FA specific regulation. In turn, the induction of growth defects in the diverse environmental conditions shared by UFA and DUFA overproduction are likely not related to regulation or increased energy demand, but instead to changes in membrane permeability and structure. *E. coli* has been shown to be able to survive on a wide range of UFA as a percentage of total FAs(3, 4). It remains to be seen whether WT *E. coli* strains with higher natural levels of UFA or CFA would better tolerate overproduction of either of these FAs.

Besides the above-mentioned modified FAs in phospholipid membrane, some Gram-positive bacteria can also produce internally IBFAs(6, 44). The best characterized internally IBFA is tuberculostearic acid (TSBA, 10-methylstearic acid) which is found in phospholipids (such as phosphatidylethanolamine, phosphatidylinositol, and diphosphatidylglycerol) and glycolipids (such as phosphatidylinositol mannosides and lipoarabinomannans) in *Mycobacterium tuberculosis* and related species(75, 76). Though the biosynthesis mechanism is still not fully understood, it was speculated that TSBA is synthesized from oleic acid by a two-step process of

methylenation and reduction, which is catalyzed by an S-adenosyl-L-methionine (SAM)-dependent methyltransferase (BfaB, encoded by *bfaB*) and a FAD-binding oxidoreductase (BfaA, encoded by *bfaA*), respectively(6, 44). However, these two enzyme's direct substrates are still not confirmed. Our analysis of the LC-MS/MS results detected IBFA only in PG, PE, and PS; this suggests that the substrate of the IBFA pathway are likely PS, PE, and PG, but not PA, CDP-DAG, or CL. The significantly higher observed IBFA content in PG compared to PE and PS is likely caused by the dynamic distribution of phospholipids between the inner and outer membrane(77). As the IBFA synthesizing proteins BfaA and BfaB were not predicted to contain any predicted secretion tags or transmembrane sequences (Figure 3.18S), the IBFA pathway most likely methylates phospholipid FAs in the inner membrane, which then diffuse to the outer membrane. Because PG is enriched in the inner membrane, limits in phospholipid transfer between two membranes can cause a higher IBFA content in PG than that in PS or PE(77). The almost complete conversion of PG from UFA to IBFA at the sn1 position of phospholipids suggests that the IBFA pathway prefers ω7-UFAs at the sn-1. Integration of the IBFA pathway incurred little to no metabolic burden, and the expression profile did not differ from UFA, or CFA, indicative that the *E. coli* chassis was able to integrate a novel phospholipid species. These results also reveal the substrate specificity of IBFA pathway and the distribution of IBFAs in each phospholipid species. Such knowledge can be used to guide future engineering efforts to further tailoring and increasing IBFA content.

Coupling the phenotypic data with the transcriptomic data, we demonstrate the ability of the *E. coli* chassis to tolerate CFA and IBFA induced changes while also not incurring greater metabolic costs. Synthesis of the novel DUFA-enriched strain however, resulted in an altered transcriptional state, decreased stress tolerance as compared to UFA, CFA, and IBFA, and

significant metabolic alteration in the host. This resulted in 26 highly DE transcriptional regulators unique to DUFA production, such as redox stress regulating genes *soxR/S* as well as *fur,* the iron regulation repressor, which is itself known to be regulated by *soxR/S(78)*. Production of DUFA required overexpression of ferredoxin, and we found many of the iron transport and utilization genes commonly expressed during iron-starvation conditions, such as the *ent, fec, fep,* and *fhu* operons(79), to be significantly upregulated. The increased need for iron could be redirecting resources needed for other cellular processes, resulting in activation of the *fur* iron-starvation transcriptional cascade of and the *soxR/S* stress response regulon; *fur* modulate expression of at least 81 gene targets, and indirectly regulates hundreds more(79), at the cost of significant intracellular resources. Δ*fur*/Δ*rhyB* has been shown to rescue expression of the *fur* regulated *sdh* TCA cycle genes and *sodB(80)* while also increasing expression of iron uptake and acquisition systems in the presence of iron, and is a possible target for further genetic manipulation to minimize the trade-of between DUFA production and growth.

The cell likely attempts to buffer the changes to FA composition by inducing a natural metabolic response to allow for a similar maximal growth rate during the phenotypic growth assays. Upregulation of the *sdh* operon generates the sRNA *sdhX,* and is theorized to act as a metabolic "overflow" mechanism to shunt acetyl-CoA into acetate for excretion in aerobic conditions in the presence of glucose during high TCA cycle expression(81). Increased *sdh* expression and evidence of overflow metabolism has also been reported in *E. coli* designed for increased FFA production, further defining the important relationship between acetyl-CoA partitioning and the production of FFA(82). DUFA-enrichment resulted in increased expression of the entire *sdh* operon, likely buffering the metabolic burden of double-bound FA production via the aforementioned mechanism. Overall, though growth rate was affected in some conditions, the

E. coli chassis largely tolerates DUFA-enrichment and other changes in lipid profiles, while continuing to maintain normal final cell density for all environmental conditions tested.

Future work is needed to elucidate the effects of MUFA- and FIBFA-enrichment on stress tolerance and the transcriptional state of the cell. The initial results indicate that sgRNAs are an excellent method for targeted repression of UFA biosynthesis in *E. coli*; however, more rounds of engineering may further reduce metabolic burden. Alternatively, FIBFA enrichment required the co-expression of two plasmids, and thus likely induces a greater stress on the expression profile of the cell. The use of sophisticated tools such as metabolic flux analysis could identify further refinements for reducing the effects of both increased expression and altered FA profiles on the host(83, 84).

Overall, this work details the synthetic alterations which enabled the production of altered FA profiles, tailored towards the generation of highly valuable precursor molecules, some of which (*e.g.*, MUFAs, DUFA, IBFAs) have never been produced in the *E. coli* before. We demonstrate that while some of these genetic changes do result in decreased stress tolerance, there is evidence that expression driven regulation of metabolic networks induces a transcriptional response which allows strains producing UFA, CFA, and IBFA to exhibit a similar expression profile. In addition to creating these novel strains, our work also identified the substrates for the *M. smegmatis* IBFA biosynthesis pathway. This represents a substantial increase in knowledge concerning the effects of FA composition on the bacterial cell, and on the effective generation of high FA-titer producing *E. coli*.

3.5 Acknowledgements

This work was created in collaboration with Wenqin Bai, Christopher Hartline, Shaojie Wang, Bin Wang, Jie Ning, Fong-Fu Hsu, Gautam Dantas, and Fuzhong Zhang, *In Review*. F.Z., W.B.,

W.E.A., and G.D. conceived the project. W.B. and S.W. engineered the microbial strains, performed fermentation, and FA analysis; F.H. performed the lipid analysis; C.H. performed phenotypic profiling and growth experiments; W.E.A. performed the transcriptomic analysis. All authors analyzed the data and wrote the paper. This work was supported by the Department of Energy (DESC0018324).

3.5 References

1. Zhang Y, Rock CO. Membrane lipid homeostasis in bacteria. Nature Reviews Microbiology. 2008;6(3):222-33.

2. Chang YY, Cronan JE. Membrane cyclopropane fatty acid content is a major factor in acid resistance of Escherichia coli. Molecular Microbiology. 1999;33(2):249-59.

3. Sinensky M. Homeoviscous adaptation--a homeostatic process that regulates the viscosity of membrane lipids in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America. 1974;71(2):522-5.

4. Budin I, De Rond T, Chen Y, Chan LJG, Petzold CJ, Keasling JD. Viscous control of cellular respiration by membrane lipid composition. Science. 2018;362(6419):1186-9.

5. Hildebrand JG, Law JH. Fatty Acid Distribution in Bacterial Phospholipids. The Specificity of the Cyclopropane Synthetase Reaction. Biochemistry. 1964;3:1304-8.

6. Bai W, Geng W, Wang S, Zhang F. Biosynthesis, regulation, and engineering of microbially produced branched biofuels. Biotechnol Biofuels. 2019;12:84.

7. Kim HM, Chae TU, Choi SY, Kim W, Lee SY. Engineering of an oleaginous bacterium for the production of fatty acids and fuels. Nature Chemical Biology. 2019;15(7):721-9.

8. Abraham, J., Domb, Raphael, Nudelman. Biodegradable polymers derived from natural fatty acids. Journal of Polymer Science Part A: Polymer Chemistry. 1995.

9. Akhtar MK, Turner NJ, Jones PR. Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. Pnas. 2013;110(1):87-92.10.

Carroll AL, Desai SH, Atsumi S. Microbial production of scent and flavor compounds. Curr Opin Biotechnol. 2016;37:8-15.

11. Jiang W, Qiao JB, Bentley GJ, Liu D, Zhang F. Modular pathway engineering for the microbial production of branched-chain fatty alcohols. Biotechnol Biofuels. 2017;10:244.

12. Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, et al. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature. 2010;463(7280):559-62.

13. Yan Q, Pfleger BF. Revisiting metabolic engineering strategies for microbial synthesis of oleochemicals. Metab Eng. 2020;58:35-46.

14. Peralta-Yahya PP, Zhang F, del Cardayre SB, Keasling JD. Microbial engineering for the production of advanced biofuels. Nature. 2012;488(7411):320-8.

15. Lee SK, Chou H, Ham TS, Lee TS, Keasling JD. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. Curr Opin Biotechnol. 2008;19(6):556-63.

16. Tao H, Guo D, Zhang Y, Deng Z, Liu T. Metabolic engineering of microbes for branched-chain biodiesel production with low-temperature property. Biotechnol Biofuels. 2015;8:92.

17. Cronan JE, Jr. Phospholipid modifications in bacteria. Curr Opin Microbiol. 2002;5(2):202-5.

18. Willdigg JR, Helmann JD. Mini Review: Bacterial Membrane Composition and Its Modulation in Response to Stress. Front Mol Biosci. 2021;8:634438-.

19. Tan Z, Yoon JM, Nielsen DR, Shanks JV, Jarboe LR. Membrane engineering via trans unsaturated fatty acids production improves Escherichia coli robustness and production of biorenewables. Metab Eng. 2016;35:105-13.

20. Dunlop MJ. Engineering microbes for tolerance to next-generation biofuels. Biotechnology for Biofuels. 2011;4(1):32.

21. Schirmer A, Rude M, Li X, Popova E, Cardayre SBD. Microbial Biosynthesis of Alkanes. Science. 2010;329(5991):559-62.

22. Goh EB, Baidoo EE, Keasling JD, Beller HR. Engineering of bacterial methyl ketone synthesis for biofuels. Appl Environ Microbiol. 2012;78(1):70-80.

23. Nawabi P, Bauer S, Kyrpides N, Lykidis A. Engineering Escherichia coli for biodiesel production utilizing a bacterial fatty acid methyltransferase. Appl Environ Microbiol. 2011;77(22):8052-61.

24. Bowen CH, Bonin J, Kogler A, Barba-Ostria C, Zhang F. Engineering Escherichia coli for Conversion of Glucose to Medium-Chain omega-Hydroxy Fatty Acids and alpha,omega-Dicarboxylic Acids. ACS Synth Biol. 2016;5(3):200-6.

25. Jiang W, Jiang Y, Bentley GJ, Liu D, Xiao Y, Zhang F. Enhanced production of branched-chain fatty acids by replacing beta-ketoacyl-(acyl-carrier-protein) synthase III (FabH). Biotechnol Bioeng. 2015;112(8):1613-22.

26. Bentley GJ, Jiang W, Guaman LP, Xiao Y, Zhang F. Engineering Escherichia coli to produce branched-chain fatty acids in high percentages. Metab Eng. 2016;38:148-58.

27. Lennen RM, Pfleger BF. Modulating Membrane Composition Alters Free Fatty Acid Tolerance in Escherichia coli. PLOS ONE. 2013;8(1):e54031.

28. Grogan DW, Cronan JE. Cyclopropane ring formation in membrane lipids of bacteria. Microbiology and Molecular Biology Reviews. 1997;61(4):429-41.

29. Rowlett VW, Mallampalli VKPS, Karlstaedt A, Dowhan W, Taegtmeyer H, Margolin W, et al. Impact of Membrane Phospholipid Alterations in Escherichia coli on Cellular Function and Bacterial Stress Adaptation. Journal of Bacteriology. 2017;199(13).

30. Lennen RM, Kruziki MA, Kumar K, Zinkel RA, Burnum KE, Lipton MS, et al. Membrane stresses induced by overproduction of free fatty acids in Escherichia coli. Applied and environmental microbiology. 2011;77(22):8114-28.

31. Royce LA, Liu P, Stebbins MJ, Hanson BC, Jarboe LR. The damaging effects of short chain fatty acids on Escherichia coli membranes. Appl Microbiol Biotechnol. 2013;97(18):8317-27.

32. Hobby CR, Herndon JL, Morrow CA, Peters RE, Symes SJK, Giles DK. Exogenous fatty acids alter phospholipid composition, membrane permeability, capacity for biofilm formation, and antimicrobial peptide susceptibility in Klebsiella pneumoniae. Microbiologyopen. 2019;8(2):e00635.

33. Cao Y, Yang J, Xian M, Xu X, Liu W. Increasing unsaturated fatty acid contents in Escherichia coli by coexpression of three different genes. Appl Microbiol Biotechnol. 2010;87(1):271-80.

34. Cao Y, Liu W, Xu X, Zhang H, Wang J, Xian M. Production of free monounsaturated fatty acids by metabolically engineered Escherichia coli. Biotechnology for Biofuels. 2014;7(1):59.

35. Wolff RL, Christie WW, Pédrono F, Marpeau AM, Tsevegsüren N, Aitzetmüller K, et al. Δ5-Olefinic acids in the seed lipids from fourEphedraspecies and their distribution between the α and β positions of triacylglycerols. Characteristics common to coniferophytes and cycadophytes. Lipids. 1999;34(8):855-64.

36. Marr AG, Ingraham JL. EFFECT OF TEMPERATURE ON THE COMPOSITION OF FATTY ACIDS IN ESCHERICHIA COLI. Journal of bacteriology. 1962;84(6):1260-7.

37. Cronan JE, Gelmann EP. An Estimate of the Minimum Amount of Unsaturated Fatty Acid Required for Growth of Escherichia coli. Journal of Biological Chemistry. 1973;248(4):1188-95.

38. Shuntaro M, Bakku RK, Iwane S. Expression of Genes for a Flavin Adenine Dinucleotide-Binding Oxidoreductase and a Methyltransferase from Mycobacterium

chlorophenolicum Is Necessary for Biosynthesis of 10-Methyl Stearic Acid from Oleic Acid in Escherichia coli. Frontiers in Microbiology. 2017;8.

39. Taylor FR, Grogan DW, Cronan JE, Jr. Cyclopropane fatty acid synthase from Escherichia coli. Methods Enzymol. 1981;71 Pt C:133-9.

40. Zhang F, Ouellet M, Batth TS, Adams PD, Petzold CJ, Mukhopadhyay A, et al. Enhancing fatty acid production by the expression of the regulatory transcription factor FadR. Metabolic Engineering. 2012;14(6):653-60.

41. Shabala L, Ross T. Cyclopropane fatty acids improve Escherichia coli survival in acidified minimal media by reducing membrane permeability to H+ and enhanced ability to extrude H+. Res Microbiol. 2008;159(6):458-61.

42. Cronan Jr JE, Nunn WD, Batchelor JG. Studies on the biosynthesis of cyclopropane fatty acids in Escherichia coli. Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism. 1974;348(1):63-75.

43. Cronan JE, Jr. Phospholipid alterations during growth of Escherichia coli. Journal of bacteriology. 1968;95(6):2054-61.

44. Machida S, Bakku RK, Suzuki I. Expression of Genes for a Flavin Adenine Dinucleotide-Binding Oxidoreductase and a Methyltransferase from Mycobacterium chlorophenolicum Is Necessary for Biosynthesis of 10-Methyl Stearic Acid from Oleic Acid in Escherichia coli. Front Microbiol. 2017;8:2061.

45. Richard D, Kefi K, Barbe U, Bausero P, Visioli F. Polyunsaturated fatty acids as antioxidants. Pharmacological Research. 2008;57(6):451-5.

46. Chazarretacifre L, Martiarena L, Mendoza DD, Altabe SG. Role of Ferredoxin and Flavodoxins in Bacillus subtilis Fatty Acid Desaturation . Journal of Bacteriology. 2011;193(16):4043-8.

47. Bonamore A, Macone A, Colotti G, Matarese RM, Boffi A. The desaturase from Bacillus subtilis, a promising tool for the selective olefination of phospholipids. Journal of Biotechnology. 2006;121(1):49-53.

48. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell. 2013;152(5):1173-83.

49. Hawrot E, Kennedy EP. Phospholipid composition and membrane function in phosphatidylserine decarboxylase mutants of Escherichia coli. Journal of Biological Chemistry. 1978;253(22):8213-20.

50. Parsons JB, Rock CO. Bacterial lipids: Metabolism and membrane homeostasis. Progress in Lipid Research. 2013;52(3):249-76.

51. Altabe SG, Aguilar P, Caballero GM, Mendoza DD. The Bacillus subtilis Acyl Lipid Desaturase Is a Δ5 Desaturase. Journal of Bacteriology. 2003;185(10):3228-31.

52. Feng Y, Cronan JE. Escherichia coli unsaturated fatty acid synthesis: complex transcription of the fabA gene and in vivo identification of the essential reaction catalyzed by FabB. J Biol Chem. 2009;284(43):29526-35.

53. Falls KC, Williams AL, Bryksin AV, Matsumura I. Escherichia coli deletion mutants illuminate trade-offs between growth rate and flux through a foreign anabolic pathway. PloS one. 2014;9(2):e88159-e.

54. Oyarzún DA, Stan G-BV. Synthetic gene circuits for metabolic control: design trade-offs and constraints. Journal of The Royal Society Interface. 2013;10(78):20120671.

55. Dunlop MJ, Keasling JD, Mukhopadhyay A. A model for improving microbial biofuel production using a synthetic feedback loop. Syst Synth Biol. 2010;4(2):95-104.

56. Jiang W, Gu P, Zhang F. Steps towards 'drop-in' biofuels: focusing on metabolic pathways. Current Opinion in Biotechnology. 2018;53:26-32.

57. Ramrakhiani L, Chand S. Recent progress on phospholipases: different sources, assay methods, industrial potential and pathogenicity. Appl Biochem Biotechnol. 2011;164(7):991-1022.

58. Karasawa K, Nojima S. Lysophospholipases from Escherichia coli. Methods in Enzymology. 1991;197:437-45.

59. Masayama A, Kato S, Terashima T, Molgaard A, Hemmi H, Yoshimura T, et al. Bacillus subtilis Spore Coat Protein LipC Is a Phospholipase B. Bioscience, Biotechnology, and Biochemistry. 2010;74(1):24-30.

60. Hari SB, Grant RA, Sauer RT. Structural and Functional Analysis of E. coli Cyclopropane Fatty Acid Synthase. Structure. 2018;26(9):1251-8 e3.

61. Hsu L, Jackowski S, Rock CO. Isolation and characterization of Escherichia coli K-12 mutants lacking both 2-acyl-glycerophosphoethanolamine acyltransferase and acyl-acyl carrier protein synthetase activity. J Biol Chem. 1991;266(21):13783-8.

62. Jackowski S, Rock CO. Transfer of fatty acids from the 1-position of phosphatidylethanolamine to the major outer membrane lipoprotein of Escherichia coli. J Biol Chem. 1986;261(24):11328-33.

63. Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. PLoS One. 2008;3(11):e3647.

64. Jiang Y, Chen B, Duan C, Sun B, Yang J, Yang S. Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. Appl Environ Microbiol. 2015;81(7):2506-14.

65. Folch JMS, M.M L, Stanley GHS. A simple method for the isolation and purification of total lipids from animal Tissues. Journal of the Science of Food & Agriculture. 1957;22(1):24-36.

66. Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. Can J Biochem Physiol. 1959;37(8):911-7.

67. Hsu F-F, Turk J. Electrospray ionization with low-energy collisionally activated dissociation tandem mass spectrometry of glycerophospholipids: Mechanisms of fragmentation and structural characterization. Journal of Chromatography B. 2009;877(26):2673-95.

68. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. PLOS ONE. 2015;10(5):e0128036.

69. Henson WR, Campbell T, DeLorenzo DM, Gao Y, Berla B, Kim SJ, et al. Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. Metabolic engineering. 2018;49:69-83.

70. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357-9.

71. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.

72. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550-.

73. Dixon P. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science. 2003;14(6):927-30.

74. Santoscoy MC, Jarboe LR. A systematic framework for using membrane metrics for strain engineering. Metabolic Engineering. 2021;66:98-113.

75. Lennarz WJ, Scheuerbrandt G, Bloch K. The biosynthesis of oleic and 10-methylstearic acids in Mycobacterium phlei. Journal of Biological Chemistry. 1962;237(3):664-71.

76. Bansalmutalik R, Nikaido H. Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(13):4958-63.

77. Bayer MH, Bayer ME. Phosphoglycerides and phospholipase C in membrane fractions of Escherichia coli B. Journal of Bacteriology. 1985;162(1):50-4.

78. Hantke K. Iron and metal regulation in bacteria. Current Opinion in Microbiology. 2001;4(2):172-7.

79. Seo SW, Kim D, Latif H, O'Brien EJ, Szubin R, Palsson BO. Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in Escherichia coli. Nat Commun. 2014;5:4910-.

80. Massé E, Gottesman S. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. Proceedings of the National Academy of Sciences. 2002;99(7):4620-5.

81. De Mets F, Van Melderen L, Gottesman S. Regulation of acetate metabolism and coordination with the TCA cycle via a processed small RNA. Proceedings of the National Academy of Sciences. 2019;116(3):1043.

82. Youngquist JT, Korosh TC, Pfleger BF. Functional genomics analysis of free fatty acid production under continuous phosphate limiting conditions. J Ind Microbiol Biotechnol. 2017;44(4-5):759-72.

83. Ando D, Garcia Martin H. Two-Scale 13C Metabolic Flux Analysis for Metabolic Engineering. In: Jensen MK, Keasling JD, editors. Synthetic Metabolic Pathways: Methods and Protocols. New York, NY: Springer New York; 2018. p. 333-52.

84. He L, Xiao Y, Gebreselassie N, Zhang F, Antoniewiez MR, Tang YJ, et al. Central metabolic responses to the overproduction of fatty acids in Escherichia coli based on 13C-metabolic flux analysis. Biotechnology and bioengineering. 2014;111(3):575-85.

85. Budin I, de Rond T, Chen Y, Chan Leanne Jade G, Petzold Christopher J, Keasling Jay D. Viscous control of cellular respiration by membrane lipid composition. Science. 2018;362(6419):1186-9.

86. Chen YY, Gänzle MG. Influence of cyclopropane fatty acids on heat, high pressure, acid and oxidative resistance in Escherichia coli. International Journal of Food Microbiology. 2016;222:16-22.

87. Luo LH, Seo P-S, Seo J-W, Heo S-Y, Kim D-H, Kim CH. Improved ethanol tolerance in Escherichia coli by changing the cellular fatty acids composition through genetic manipulation. Biotechnology Letters. 2009;31(12):1867.

88. Nielsen H. Predicting Secretory Proteins with SignalP. In: Kihara D, editor. Protein Function Prediction: Methods and Protocols. New York, NY: Springer New York; 2017. p. 59-73.

89. Hofmann K, Stoffel W. TMbase-A database of membrane spanning proteins segments. Biol Chem Hoppe-Seyler. 1993;374:166.

3.6 Appendix (Supplemental Material)

3.6.1 Tables

Table 3. 2 The strains used in this study

Strains	Genotype	Description
E. coli DH10β	F^- endA1 recA1 galE15 galK16 nupG rpsLΔlacX74 Φ80lacZΔM15 araD139Δ(ara, leu) 7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ^-	Host cell for gene cloning
E. coli MG1655	K-12 F^- λ^- *ilvG*$^-$ *rfb-50 rph-1*	Host cell for fatty acids (FA) production
WT	MG1655ΔfadE	Host cell (with a *fadE* gene deletion) for FA production. Gene *fadE* encodes an acyl-CoA dehydrogenase, which is involved in FFA degradation (β-oxidation).
ω7-UFA-1	MG1655ΔfadEΔcfa	ω7-UFA producing strain with *fadE* and *cfa* gene deletions. Gene *cfa* encodes a cyclopropane fatty acids (CFA) synthase, which is involved in CFA synthesis.
ω7-UFA-2	MG1655ΔfadEΔcfa/pE8a-fadR	ω7-UFA producing strain harboring *fadR* encodes transcription activator FadR, which is involved in FA production.
ω7-UFA-3	MG1655ΔfadEΔcfa/pE8a-fadR, pB6c-fabB	ω7-UFA producing strain harboring *fabB* encodes β-ketoacyl-ACP synthase, which is involved in unsaturated fatty acids (UFA) biosynthesis.
ω7-UFA-4	MG1655ΔfadEΔcfa/pE8a-fadR, pB6c-fabAB	ω7-UFA producing strain harboring *fabA* and *fabB,* which are involved in UFA biosynthesis.
ω7-UFA-5	MG1655ΔfadEΔcfaΔfabR/pE8a-fadR	ω7-UFA producing strain with *fadE*, *cfa* and *fabR* gene deletions and harboring *fadR* gene. Gene *fabR* encodes a transcription repressor FabR, which controls UFA synthesis.
CFA-1	MG1655ΔfadE/pE8a-cfa(Ec)	CFA producing strain harboring *cfa* (Ec) gene, which is from *Escherichia coli*.
CFA-2	MG1655ΔfadEΔcfaΔfabR/pE8a-cfa(Ec)-fadR	CFA producing strain with *fadE*, *cfa* and *fabR* gene deletions, and harboring *cfa* (Ec) and *fadR* genes.
CFA-3	MG1655ΔfadEΔcfaΔfabR/pE8a-cfa(Cb)-fadR	CFA producing strain with *fadE*, *cfa* and *fabR* gene deletions, and harboring *cfa* (Cb) and *fadR* genes. *cfa* (Cb) is from *Clostridium butyricum*.

CFA-4	MG1655ΔfadEΔcfaΔfabR/pE8a-cfa(Ec)-cfa(Cb)-fadR	CFA producing strain with *fadE*, *cfa* and *fabR* gene deletions, and harboring *cfa* (Ec) and *cfa* (Cb) and *fadR* genes.
IBFA-1	MG1655ΔfadE/pE8a-BfaAB(Ro)	BCFA producing strain with *fadE* gene deletion and harboring *bfaA* and *bfaB* genes, which are responsible for BCFA biosynthesis and from *Rhodococcus opacus* PD630.
IBFA-2	MG1655ΔfadEΔcfa/pE8a-BfaAB(Ro)	BCFA producing strain with *fadE* and *cfa* gene deletions and harboring *bfaA* and *bfaB* genes.
IBFA-3	MG1655ΔfadEΔcfa/pE8a-BfaAB(Ro)-fadR	BCFA producing strain with *fadE* and *cfa* gene deletions and harboring *bfaA*, *bfaB*, and *fadR* genes.
IBFA-4	MG1655ΔfadEΔcfaΔfabR/pE8a-BfaAB(Ro)-fadR	BCFA producing strain with *fadE*, *cfa,* and *fabR* gene deletions and harboring *bfaA*, *bfaB*, and *fadR* genes.
IBFA-5	MG1655ΔfadEΔcfaΔfabR/pE8a-BfaAB(Ms)-fadR	BCFA producing strain with *fadE*, *cfa,* and *fabR* gene deletions and harboring *bfaA*, *bfaB*, and *fadR* genes. bfaAB(Ms) are from *Mycobacterium smegmatis*.
IBFA-6	MG1655ΔfadEΔcfaΔfabR/pE8a-BfaAB(Tc)-fadR	BCFA producing strain with *fadE*, *cfa,* and *fabR* gene deletions and harboring *bfaA*, *bfaB*, and *fadR* genes. bfaAB(Tc) are from *Thermomonospora curvata*.
DUFA-1	MG1655ΔfadE/pE8a-Des5(Bs)	DUFA producing strain with *fadE* gene deletion and harboring *des5*(Bs) gene encodes a Δ5 acyl lipid desaturase, which synthesize UFA with a double bond at position Δ5. Des5(Bs) is from *Bacillus subtilis*.
DUFA-2	MG1655ΔfadEΔcfaΔfabR/pE8a-Des5(Bs)-fadR	DUFA producing strain with *fadE, cfa, and fabR* gene deletions and harboring *des5*(Bs) and *fadR* gene.
DUFA-3	MG1655ΔfadEΔcfaΔfabR/pE8a-Des5(Bs)-Fd(Ec)-fadR	DUFA producing strain with *fadE, cfa, and fabR* gene deletions and harboring *des5*(Bs), *fd*(Ec) and *fadR* gene. *Fd*(Ec) encodes ferredoxin from *E. coli*.
DUFA-4	MG1655ΔfadEΔcfaΔfabR/pE8a-Des5(Bs)-Fd(Bs)-fadR	DUFA producing strain with *fadE, cfa, and fabR* gene deletions and harboring *des5*(Bs), *fd*(Bs) and *fadR* gene. *Fd*(Bs) encodes ferredoxin from *B. subtilis*.
DUFA-5	MG1655ΔfadEΔcfaΔfabR/pE8a-Des5(Bs)-Fd(An)-fadR	DUFA producing strain with *fadE, cfa, and fabR* gene deletions and harboring *des5*(Bs), *fd*(An) and *fadR* gene. *Fd*(An) encodes ferredoxin from *Anabaena* sp. 7120.
WT2	MG1655ΔfadEΔcfa/pdCas9	Control strain for UFA production with *fadE* and *cfa* gene deletions, harboring *dCas9* gene.
FabA-P1	MG1655ΔfadEΔcfa/pdCas9, pTargetF-FabA-P1	ω7-UFA inhibited-producing strain with *fadE* and *cfa* gene deletions, harboring *dCas9* gene and *fabA*-P1, which encodes small guide RNA in the promoter region of *fabA* gene.

FabA-NT1	MG1655ΔfadEΔcfa/pdCas9, pTargetF-FabA-NT1	ω7-UFA inhibited-producing strain with *fadE* and *cfa* gene deletions, harboring *dCas9* gene and *fabA*-NT1, which encodes small guide RNA in the upstream region of *fabA* gene.
FabA-NT2	MG1655ΔfadEΔcfa/pdCas9, pTargetF-FabA-NT2	ω7-UFA inhibited-producing strain with *fadE* and *cfa* gene deletions, harboring *dCas9* gene and *fabA*-NT2, which encodes small guide RNA in the middle region of *fabA* gene.
FabA-NT3	MG1655ΔfadEΔcfa/pdCas9, pTargetF-FabA-NT3	ω7-UFA inhibited-producing strain with *fadE* and *cfa* gene deletions, harboring *dCas9* gene and *fabA*-NT3, which encodes small guide RNA in the downstream of *fabA* gene.
FabB-P1	MG1655ΔfadEΔcfa/pdCas9, pTargetF-FabB-P1	ω7-UFA inhibited-producing strain with *fadE* and *cfa* gene deletions, harboring *dCas9* gene and *fabB*-P1, which encodes small guide RNA in the promoter region of *fabB* gene.
FabB-NT1	MG1655ΔfadEΔcfa/pdCas9, pTargetF-FabB-NT1	ω7-UFA inhibited-producing strain with *fadE* and *cfa* gene deletions, harboring *dCas9* gene and *fabB*-NT1, which encodes small guide RNA in the upstream region of *fabB* gene.
FabB-NT2	MG1655ΔfadEΔcfa/pdCas9, pTargetF-FabB-NT2	ω7-UFA inhibited-producing strain with *fadE* and *cfa* gene deletions, harboring *dCas9* gene and *fabB*-NT2, which encodes small guide RNA in the middle region of *fabB* gene.
FabB-NT3	MG1655ΔfadEΔcfa/pdCas9, pTargetF-FabB-NT3	ω7-UFA inhibited-producing strain with *fadE* and *cfa* gene deletions, harboring *dCas9* gene and *fabB*-NT3, which encodes small guide RNA in the downstream of *fabB* gene.
Δ5-MUFA-1	MG1655ΔfadE/pE8a-Des5(Bs)	Δ5-MUFA producing strain with *fadE* gene deletion and harboring *des5*(Bs) gene.
Δ5-MUFA-2	MG1655ΔfadEΔcfa/pE8a-Des5(Bs), pdCas9	Δ5-MUFA producing strain with *fadE* and *cfa* gene deletions and harboring *des5*(Bs) and *dCas9* genes. *dCas9* encodes a catalytically dead Cas9 mutant.
Δ5-MUFA-3	MG1655ΔfadEΔcfa/pE8a-Des5(Bs)-pTarget FabB-NT2, pdCas9	Δ5-MUFA producing strain with *fadE* and *cfa* gene deletions and harboring *des5*(Bs), *fabB-NT2,* and *dCas9* genes. *FabB*-NT2 encodes small guide RNA in the middle of *fabB* gene coding region.
Δ5-MUFA-4	MG1655ΔfadEΔcfa/pE8a-Des5(Bs)-Fd (Ec)-pTarget FabB-NT2, pdCas9	Δ5-MUFA producing strain with *fadE* and *cfa* gene deletions and harboring *des5*(Bs), *fd(Ec), fabB*-NT2, and *dCas9* genes.

Δ5-MUFA-5	MG1655ΔfadEΔcfa/ pE8a-Des5(Bs)-Fd(Bs)-pTarget FabB-NT2, pdCas9	Δ5-MUFA producing strain with *fadE* and *cfa* gene deletions and harboring *des5*(Bs), *fd(Bs), fabB*-NT2, and *dCas9* genes.
Δ5-MUFA-6	MG1655ΔfadEΔcfa/ pE8a-Des5(Bs)-Fd(An)-pTarget FabB-NT2, pdCas9	Δ5-MUFA producing strain with *fadE* and *cfa* gene deletions and harboring *des5*(Bs), *fd(An), fabB*-NT2, and *dCas9* genes.
FIBFA-1	MG1655ΔfadEΔcfaΔfabR/pE8a-BfaAB(Ms)-fadR	Negative control strain for production of free branched chain fatty acids (FBCFA). It has *fadE, cfa* and fabR gene deletions and harbors *bfaAB(Ms)* and *fadR* genes.
FIBFA-2	MG1655ΔfadEΔcfaΔfabR/pE8a-BfaAB(Ms)-fadR, pB8k-PldA(Ec)	FBCFA producing strain with *fadE*, *cfa,* and *fabR* gene deletions and harboring bfaAB(Ms), *fadR*, and *pldA(Ec)* genes. *PldA(Ec)* encodes outer membrane phospholipase A, which is from *E. coli*.
FIBFA-3	MG1655ΔfadEΔcfaΔfabR/pE8a-BfaAB(Ms)-fadR, pB8k-PldB(Ec)	FBCFA producing strain with *fadE*, *cfa,* and *fabR* gene deletions and harboring *bfaAB*(Ms), *fadR*, and *pldB(Ec)* genes. *PldB(Ec)* encodes inner membrane phospholipase B, which is from *E. coli*.
FIBFA-4	MG1655ΔfadEΔcfaΔfabR/pE8a-BfaAB(Ms)-fadR, pB8k-LipC(Bs)	FBCFA producing strain with *fadE*, *cfa,* and *fabR* gene deletions and harboring *bfaAB*(Ms), *fadR*, and *lipC(Bs)* genes. *LipC(Bs)* encodes cytoplasm phospholipase, which is from *B. subtilis.*
FIBFA-5	MG1655ΔfadEΔcfaΔfabR/pE8a-BfaAB(Ms)-fadR, pB8k-PldA(Ec)-PldB(Ec)	FBCFA producing strain with *fadE*, *cfa,* and *fabR* gene deletions and harboring *bfaAB*(Ms), *fadR*, *pldA(Ec),* and *pldB(Ec)* genes.
FIBFA-6	MG1655ΔfadEΔcfaΔfabR/pE8a-BfaAB(Ms)-fadR, pB8k-PldA(Ec)-LipC(Bs)	FBCFA producing strain with *fadE*, *cfa,* and *fabR* gene deletions and harboring *bfaAB*(Ms), *fadR*, *pldA(Ec),* and *lipC(Bs)* genes.
FIBFA-7	MG1655ΔfadEΔcfaΔfabRΔaas/ pE8a-BfaAB(Ms)-fadR, pB8k-PldA(Ec)	FBCFA producing strain with *fadE*, *cfa, fabR,* and *aas* gene deletions and harboring *bfaAB*(Ms), *fadR*, and *pldA(Ec)* genes. *Aas* gene encodes bifunctional acyl-[acyl carrier protein] synthetase/2-acylglycerophosphoethanolamine acyltransferase, which is responsible for the regeneration of phosphatidylethanolamine from free FAs.
FIBFA-8	MG1655ΔfadEΔcfaΔfabRΔaas ΔfadD/pE8a-BfaAB(Ms)-fadR, pB8k-PldA(Ec)	FBCFA producing strain with *fadE*, *cfa, fabR, aas*, and *fadD* gene deletions and harboring *bfaAB*(Ms), *fadR*, and *pldA(Ec)*

		genes. *FadD* gene encodes fatty acyl-coA synthetase, which activates free FAs to acyl-coA.

Table 3. 3S Plasmids used in the study

Plasmids	Descriptions (Antibiotic, replicate origin, promoter, expressing genes, (regulator))
pE8a-fadR	FadR expressing plasmid (Amp^R, ColE1, P_{BAD}, *fadR*, (AraC))
pB6c-fabB	FabB expressing plasmid (Cm^R, pBBR1, $P_{LlacO-1}$, *fabB*, (LacI))
pB6c-fabAB	FabA and FabB expressing plasmid (Cm^R, pBBR1, $P_{LlacO-1}$, *fabA* and *fabB*, (LacI))
pE8a-cfa(Ec)	CFA producing plasmid (Amp^R, ColE1, P_{BAD}, *cfa* from *E. coli*, (AraC))
pE8a-cfa(Ec)-fadR	CFA producing plasmid (Amp^R, ColE1, P_{BAD}, *cfa* from *E. coli* and *fadR*, (AraC))
pE8a-cfa(Cb)-fadR	CFA producing plasmid (Amp^R, ColE1, P_{BAD}, *cfa* from *Clostridium butyricum* and *fadR*, (AraC))
pE8a-cfa(Ec)-cfa(Cb)-fadR	CFA producing plasmid (Amp^R, ColE1, P_{BAD}, *cfa* from *E. coli* and *Clostridium butyricum* and *fadR*, (AraC))
pE8a-BfaAB(Ro)	BCFA producing plasmid (Amp^R, ColE1, P_{BAD}, *BfaA and BfaB* from *Rhodococcus. opacus*, (AraC))
pE8a-BfaAB(Ro)-fadR	BCFA producing plasmid (Amp^R, ColE1, P_{BAD}, *BfaA and BfaB* from *Rhodococcus. opacus* and *fadR*, (AraC))
pE8a-BfaAB(Tc)-fadR	BCFA producing plasmid (Amp^R, ColE1, P_{BAD}, *BfaA and BfaB* from *Thermomonospora curvata* and *fadR*, (AraC))
pE8a-BfaAB(Ms)-fadR	BCFA producing plasmid (Amp^R, ColE1, P_{BAD}, *BfaA and BfaB* from *Mycolicibacterium smegmatis* and *fadR*, (AraC))
pE8a-Des5(Bs)	Δ5-MUFA producing plasmid (Amp^R, ColE1, P_{BAD}, *Des5* from *Bacillus subtilis*, (AraC))
pE8a-Des5(Bs)-fadR	Δ5-MUFA producing plasmid (Amp^R, ColE1, P_{BAD}, *Des5* from *Bacillus subtilis* and *fadR*, (AraC))
pE8a-Des5(Bs)-Fd(Ec)-fadR	Δ5-MUFA producing plasmid (Amp^R, ColE1, P_{BAD}, *Des5* from *Bacillus subtilis,* ferredoxin gene from *E. coli* and *fadR*, (AraC))
pE8a-Des5(Bs)-Fd(Bs)-fadR	Δ5-MUFA producing plasmid (Amp^R, ColE1, P_{BAD}, *Des5* and ferredoxin gene from *Bacillus subtilis* and *fadR*, (AraC))
pE8a-Des5(Bs)-Fd(An)-fadR	Δ5-MUFA producing plasmid (Amp^R, ColE1, P_{BAD}, *Des5* from *Bacillus subtilis*, ferredoxin gene from *Anabaena* sp. 7120 and *fadR*, (AraC))
pCas9	dCas9 expressing plasmid (Kan^R, pSC101, constitutive promoter , *Cas9*, None)
pTargetF-cfa	sgRNA expressing plasmid ($Strep^R$, ColE1, J23119(SpeI) promoter, sgRNA and homologous sequence of *cfa* gene, None)
pTargetF-fabR	sgRNA expressing plasmid ($Strep^R$, ColE1, J23119(SpeI) promoter, sgRNA and homologous sequence of *fabR* gene, None)

pTargetF-aas	sgRNA expressing plasmid ($Strep^R$, ColE1, J23119(SpeI) promoter, sgRNA and homologous sequence of *aas* gene, None)
pTargetF-fadD	sgRNA expressing plasmid ($Strep^R$, ColE1, J23119(SpeI) promoter, sgRNA and homologous sequence of *fadD* gene, None)
pdCas9	dCas9 expressing plasmid (Cm^R, p15A, $P_{tetR/tetA}$, *dCas9*, (TetR))
pTargetF-FabA-P1	sgRNA expressing plasmid ($Strep^R$, ColE1, J23119(SpeI) promoter, sgRNA in the promoter region of *fabA* gene, None)
pTargetF-FabA-NT1	sgRNA expressing plasmid ($Strep^R$, ColE1, J23119(SpeI) promoter, sgRNA in the upstream of *fabA* gene coding region, None)
pTargetF-FabA-NT2	sgRNA expressing plasmid ($Strep^R$, ColE1, J23119(SpeI) promoter, sgRNA in the midstream of *fabA* gene coding region, None)
pTargetF-FabA-NT3	sgRNA expressing plasmid ($Strep^R$, ColE1, J23119(SpeI) promoter, sgRNA in the downstream of *fabA* gene coding region, None)
pTargetF-FabB-P1	sgRNA expressing plasmid ($Strep^R$, ColE1, J23119(SpeI) promoter, sgRNA in the promoter region of *fabB* gene, None)
pTargetF-FabB-NT1	sgRNA expressing plasmid ($Strep^R$, ColE1, J23119(SpeI) promoter, sgRNA in the upstream of *fabB* gene coding region, None)
pTargetF-FabB-NT2	sgRNA expressing plasmid ($Strep^R$, ColE1, J23119(SpeI) promoter, sgRNA in the middle of *fabB* gene coding region, None)
pTargetF-FabB-NT3	sgRNA expressing plasmid ($Strep^R$, ColE1, J23119(SpeI) promoter, sgRNA in the downstream of *fabB* gene coding region, None)
pE8a-Des5(Bs)-pTarget FabB-NT2	Δ5-MUFA producing plasmid (Amp^R, ColE1, P_{BAD}, *Des5* from *Bacillus subtilis*, (AraC)) and sgRNA expressing plasmid (Amp^R, ColE1, J23119(SpeI) promoter, sgRNA in the middle of *fabB* gene coding region, None)
pE8a-Des5(Bs)-Fd (Ec)-pTarget FabB-NT2	Δ5-MUFA producing plasmid (Amp^R, ColE1, P_{BAD}, *Des5* from *B. subtilis* and ferredoxin gene from *E. coli*, (AraC)) and sgRNA expressing plasmid (Amp^R, ColE1, J23119(SpeI) promoter, sgRNA in the middle of *fabB* gene coding region, None)
pE8a-Des5(Bs)-Fd(Bs)-pTarget FabB-NT2	Δ5-MUFA producing plasmid (Amp^R, ColE1, P_{BAD}, *Des5* and ferredoxin gene from *B. subtilis*, (AraC)) and sgRNA expressing plasmid (Amp^R, ColE1, J23119(SpeI) promoter, sgRNA in the middle of *fabB* gene coding region, None)
pE8a-Des5(Bs)-Fd(An)-pTarget FabB-NT2	Δ5-MUFA producing plasmid (Amp^R, ColE1, P_{BAD}, *Des5* from *B. subtilis* and ferredoxin gene from *Anabaena* sp. 7120, (AraC)) and sgRNA expressing plasmid (Amp^R, ColE1, J23119(SpeI) promoter, sgRNA in the middle of *fabB* gene coding region, None)
pB8k-PldA(Ec)	FBCFA producing plasmid (Kan^R, pBBR1, P_{BAD}, *pldA* from *E. coli*, (AraC))
pB8k-PldB(Ec)	FBCFA producing plasmid (Kan^R, pBBR1, P_{BAD}, *pldB* from *E. coli*, (AraC))

pB8k-PldB(Bs)	FBCFA producing plasmid (Kan^R, pBBR1, P_{BAD}, *pldB* from *B. subtilis*, (AraC))
pB8k-PldA(Ec)-PldB(Ec)	FBCFA producing plasmid (Kan^R, pBBR1, P_{BAD}, *pldA* and *pldB* from *E. coli*, (AraC))
pB8k-PldA(Ec)-PldB(Bs)	FBCFA producing plasmid (Kan^R, pBBR1, P_{BAD}, *pldA* from *E. coli* and *pldB* from *B. subtilis*, (AraC))

Table 3. 4S Primer sequences used in plasmid construction

Primer name	Primer sequence	Cloning fragment	Plasmid
prWB01	Ggtctcaacatatgcgtgaaggtggccg	BfaA(Ro)	pE8a-BfaAB(Ro)
prWB02	Ggtctcattaacgacgctgtactgctttgc	BfaA(Ro)	pE8a-BfaAB(Ro)
prWB03	Ggtctccttaaaggaggaataaaccatgacaa	BfaB(Ro)	pE8a-BfaAB(Ro)
prWB04	Ggtctccatccttatttccaccac	BfaB(Ro)	pE8a-BfaAB(Ro)
prWB05	Ggtctcaatgtatatctccttcttaaaag	pE8a backbone	pE8a-BfaAB(Ro)
prWB06	Ggtctccggatccaaaggatccaaactc	pE8a backbone	pE8a-BfaAB(Ro)
prWB07	Ggtctcaaggacttcgatagccaacag	fadR	pE8a-BfaAB(Ro)-fadR
prWB08	Ggtctccgctaaggatttttttatctttatcgcccctgaatggc	fadR	pE8a-BfaAB(Ro)-fadR
prWB09	Ggtctcatagtgaaaaattgcagttaaaaccgttttagagctagaaatagc	sgRNA	pTargetF-cfa
prWB10	Ggtctccattcaaaaaaagcaccgactcggtgcc	sgRNA	pTargetF-cfa
prWB11	Ggtctcagaattctatcgatatacttatacttagg	Cfa upstream fragment	pTargetF-cfa
prWB12	Ggtctcacagtcggtcacattcccacc	Cfa upstream fragment	pTargetF-cfa
prWB13	Ggtctcaactggaaaacggccttcgagtggc	Cfa downstream fragment	pTargetF-cfa
prWB14	Ggtctcaaagcttcgtgcggcgaaattcagtg	Cfa downstream fragment	pTargetF-cfa
prWB15	Ggtctccgcttagatctattaccctgttatccc	Plasmid backbone	pTargetF-cfa
prWB16	Ggtctcaactagtattatacctaggac	Plasmid backbone	pTargetF-cfa
prWB17	Ggtctcatagttgaagcgcaagccgaagcaagttttagagctagaaatagc	sgRNA	pTargetF-fabR
prWB18	Ggtctccattcaaaaaaagcaccgactcggtgcc	sgRNA	pTargetF-fabR
prWB19	Ggtctcagaattcgccaacattttgataacgc	FabR upstream fragment	pTargetF-fabR
prWB20	Ggtctcacacgatgtctgaatccttgcc	FabR upstream fragment	pTargetF-fabR
prWB21	Ggtctcacgtgatgaaacaagcaaatcaag	FabR downstream fragment	pTargetF-fabR
prWB22	Ggtctcaaagcttcgaacgtctggctctg	FabR downstream fragment	pTargetF-fabR
prWB23	Ggtctcatagtaactacactgccggggtaaagttttagagctagaaatagc	sgRNA	pTargetF-aas
prWB24	Ggtctccattcaaaaaaagcaccgactcggtgcc	sgRNA	pTargetF-aas
prWB25	Ggtctcagaattcgttacgggtgacacccaggcac	aas upstream fragment	pTargetF-aas
prWB26	Ggtctcagtgcgtaagttccgcccc	aas upstream fragment	pTargetF-aas
prWB27	Ggtctcagcacaacggctatctgcgggtgg	aas downstream fragment	pTargetF-aas
prWB28	Ggtctcaaagctttcactcatcgtgttgttccg	aas downstream fragment	pTargetF-aas

prWB29	Ggtctcctagtgctgactcaccgcaatatgcgttttagagctagaaatagcaag	sgRNA	pTargetF-fadD
prWB30	Ggtctccattcaaaaaaagcaccgactcggtgcc	sgRNA	pTargetF-fadD
prWB31	Ggtctcagaattcttgaagaaggtttggcttaaccg	fadD upstream fragment	pTargetF-fadD
prWB32	Ggtctcaccgtgccttttgccgtagatag	fadD upstream fragment	pTargetF-fadD
prWB33	Ggtctcaacggccaccaggtcaaccgggtg	fadD downstream fragment	pTargetF-fadD
prWB34	Ggtctcaaagctttcaggctttattgtccac	fadD downstream fragment	pTargetF-fadD
prWB35	Ggtctcaatgtatatctccttcttaaaag	pB6c backbone	pB6c-fabB
prWB36	Ggtctccggatccaaactcgagtaagg	pB6c backbone	pB6c-fabB
prWB37	Ggtctcaacatatgaaacgtgcagtgattactg	fabB	pB6c-fabB
prWB38	Ggtctcaatccttaatctttcagcttgcgc	fabB	pB6c-fabB
prWB39	Ggtctcactgaaggagaaattaactatgaaacgtgcagtgattac	fabB	pB6c-fabAB
prWB40	Ggtctcaatccttaatctttcagcttgcgc	fabB	pB6c-fabAB
prWB41	Ggtctccatatgagttcatcgtgtatagaagaag	cfa(Ec) fragment	pE8a-cfa(Ec)
prWB42	Ggtctcggatccttagcgagccactcgaagg	cfa(Ec) fragment	pE8a-cfa(Ec)
prWB43	Ggtctccatatgtatatctccttcttaaaagatcttttgaattc	pE8a-fadR backbone	pE8a-cfa(Cb)-fadR
prWB44	Ggtctcggatccaaactcgagtaaggatctccag	pE8a-fadR backbone	pE8a-cfa(Cb)-fadR
prWB45	Ggtctcggagaaattaactatgttgggtga	cfa(Cb) fragment	pE8a-cfa(Ec)-cfa(Cb)-fadR
prWB46	Ggtctcggatccttaaagtttatacatatattcacgcg	cfa(Cb) fragment	pE8a-cfa(Ec)-cfa(Cb)-fadR
prWB47	Ggtctccatatgactgaacaaaccattgcac	Des5(Bs) fragment	pE8a-Des5(Bs)
prWB48	Ggtctcggatcctcaggcattcttccgcagc	Des5(Bs) fragment	pE8a-Des5(Bs)
prWB49	Ggtctcggagaaattaactatgccaaagattgttatttgcc	Fd(Ec) fragment	pE8a-Des5(Bs)-Fd(Ec)-fadR
prWB50	Ggtctcggatccttaatgctcacgcgcatgg	Fd(Ec) fragment	pE8a-Des5(Bs)-Fd(Ec)-fadR
prWB51	Ggtctcggagaaattaactatggcaaagtacacaatcgtag	Fd(Bs) fragment	pE8a-Des5(Bs)-Fd(Bs)-fadR
prWB52	Ggtctcggatccctattcaaatttaagcgggtcg	Fd(Bs) fragment	pE8a-Des5(Bs)-Fd(Bs)-fadR

Table 3. 5S Phenotypic microarray log 2 fold difference of metabolic activity between FA mutants and WT measured at 590nm

590 log2 Ratio (16h)	WT	UFA-B5	CFA-B4	DUFA-B5	IBFA-B6

Negative Control	A1	0	-1.19214	0.457163	-1.25239	0.067843
Dextrin	A2	0	-0.59522	1.00073	-1.31719	-0.03772
D-Maltose	A3	0	-2.37812	-0.09796	-2.6287	-0.26094
D-Trehalose	A4	0	-1.5748	0.035988	-2.60441	-0.16874
D-cellobiose	A5	0	-1.04324	0.56612	-0.72571	0.033497
Gentiobiose	A6	0	-0.93314	0.771423	-1.21709	0.055177
Sucrose	A7	0	-1.10011	0.484857	-1.2533	-0.02921
D-Turanose	A8	0	-1.05808	0.642504	-1.26695	-0.07766
Stachyose	A9	0	-1.14762	0.404873	-1.32193	-0.14543
Positive Control	A10	0	-0.1324	0.075574	-1.34007	0.072063
pH 6	A11	0	-0.21478	0.202876	-1.44096	-0.20933
pH 5	A12	0	0.362183	0.402664	-0.05583	-0.71131
D-Raffinose	B1	0	-0.9423	0.690795	-1.2338	-0.05795
alpha-D-Lactose	B2	0	-0.73366	-0.03418	-2.38074	-0.06005
D-Melibiose	B3	0	-2.12083	0.304415	-2.54579	-0.33482
Beta-methyl-D-Glucoside	B4	0	-1.41041	0.590621	-1.57802	0.022249
D-Salicin	B5	0	-1.0608	0.693247	-1.24372	-0.01196
N-Acetyl-D-Glycosamine	B6	0	-1.14004	-0.03447	-0.20247	0.001188
N-acetyl-Beta-D-Mannosamine	B7	0	-0.90103	1.11797	-1.35867	0.081324
N-Acetyl-D-Galactosamine	B8	0	-1.23005	0.67727	-1.31188	-0.11243
N-Acetyl Neuraminic Acid	B9	0	-1.80054	0.027908	-2.42776	-0.24969
1% NaCl	B10	0	0.305591	0.62102	-0.90132	0.253019
4% NaCl	B11	0	-1.8678	0.236829	-1.1724	-0.60908

8% NaCl	B12	0	-1.25528	0.58129	-1.2241	-1.27645
alpha-D-Glucose	C1	0	0.231212	-0.20906	-0.0758	0.056939
D-Mannose	C2	0	-1.31385	0.016798	-0.89496	-0.16764
D-Fructose	C3	0	-1.42253	0.353572	-2.08625	0.024827
D-Galactose	C4	0	-1.85635	0.278265	-2.66744	-0.12163
3-Methyl Glucose	C5	0	-0.70147	1.213074	-1.01905	-0.03752
D-Fucose	C6	0	0.265407	0.862173	-0.05814	-0.08673
L-Fucose	C7	0	-1.66631	0.547906	-2.52564	-0.24304
L-Rhamnose	C8	0	-1.63717	0.702625	-2.3011	-0.10889
Inosine	C9	0	-1.90872	0.054876	-1.97409	-0.49482
1% Sodium Lactate	C10	0	0.231965	1.011034	-0.63729	-0.00052
Fusidic Acid	C11	0	-0.13514	-0.15184	-0.28816	-1.897
D-Serine	C12	0	-0.06544	-0.03071	-0.34328	-0.36166
D-Sorbitol	D1	0	-2.21256	-0.06304	-2.62906	-0.13483
D-Mannitol	D2	0	-0.57672	-0.1451	-0.32313	-0.06369
D-Arabitol	D3	0	-1.20617	0.458113	-1.15483	0.01353
myo-Inositol	D4	0	-1.13868	0.868849	-1.1194	-0.02171
Glycerol	D5	0	-1.54389	0.569279	-2	-0.05763
D-Glucose-6-PO4	D6	0	0.172488	0.3418	0.006203	-0.14629
D-Fructose-6-PO4	D7	0	-0.52413	0.455347	-1.53478	-0.47285
D-Aspartic Acid	D8	0	-1.32059	0.795574	-1.32729	-0.12099
D-Serine	D9	0	-2.0453	0.16854	-2.55043	-0.13377
Troleandomycin	D10	0	0.056322	0.393108	-1.48411	-0.53481

Rifamycin SV	D11	0	0.166129	0.291481	-1.07194	0.211453
Minocycline	D12	0	0.067253	0.296306	-0.1279	-0.21046
Gelatin	E1	0	-1.07626	0.986392	-1.2033	0.006974
Glycyl-L-Proline	E2	0	-2.69646	0.058256	-2.63888	-0.26115
L-Alanine	E3	0	-2.10835	0.090068	-2.31101	-0.16279
L-Arginine	E4	0	-0.97863	1.275032	-1.12525	0.141506
L-Aspartic Acid	E5	0	-1.81034	0.715222	-1.75094	-0.17336
L-Glutamic Acid	E6	0	-1.19564	1.136676	-1.22513	-0.04907
L-Histidine	E7	0	-0.95723	0.886078	-1.27867	-0.00506
L-Pyroglutamic Acid	E8	0	-1.18417	1.111329	-1.18847	0.049224
L-Serine	E9	0	-1.87748	0.316253	-2.42048	-0.13061
Lincomycin	E10	0	0.289613	0.662093	-0.45181	0.411768
Guanidine HCl	E11	0	0.45598	0.752559	-0.94013	0.025525
Niaproof 4	E12	0	-0.55923	-0.26364	-0.6373	-0.5995
Pectin	F1	0	-0.83855	0.611632	-0.98148	-0.05863
D-Galacturonic Acid	F2	0	-2.15265	0.110707	-2.55777	-0.36115
L-Galactonic Acid Lactone	F3	0	-2.29328	0.248076	-2.55216	-0.33941
D-Gluconinc Acid	F4	0	-0.97796	0.092793	-2.14712	-0.11049
D-Glycyronic Acid	F5	0	-1.79977	0.101093	-2.49361	-0.35287
Glucuronamide	F6	0	-0.55211	1.159661	-1.13024	-0.2592
Mucic Acid	F7	0	-2.60742	0.297775	-2.67029	-0.3142
Quinic Acid	F8	0	-1.20335	1.149082	-1.33356	0.027205
D-Saccharic Acid	F9	0	-2.00501	0.715511	-2.26363	-0.41604

Vancomycin	F10	0	-0.11069	0.160084	-0.72496	-1.68494
Tetrazolium Violet	F11	0	-0.12891	0.489375	-0.98219	0.052242
Tetrazolium Blue	F12	0	-0.03166	0.714527	-1.01346	-0.05918
p-Hydroxy-Phenylacetic Acid	G1	0	-1.28023	0.267534	-1.36307	-0.14775
Methyl Pyruvate	G2	0	-1.21398	0.017103	-1.73074	-0.33196
D-Lactic Acid Methyl Ester	G3	0	-0.62167	1.068123	-1.49434	0.188938
L-Lactic Acid	G4	0	-1.6066	0.175338	-2.58887	-0.30422
Citric Acid	G5	0	-1.55909	0.984094	-1.54749	-0.14639
alpha-Keto-Glutaric Acid	G6	0	-1.96264	0.477933	-1.99555	-0.08011
D-Malic Acid	G7	0	-2.00821	0.529688	-2.18536	-0.21415
L-Malic Acid	G8	0	-2.19556	0.042586	-2.1223	-0.29026
Bromo-Succinic Acid	G9	0	-1.62643	0.448305	-1.6978	-0.14599
Nalidixic Acid	G10	0	0.150228	-0.01096	-0.25949	-0.36101
Lithium Chloride	G11	0	-3.49027	-0.35142	-0.76046	-3.25366
Potassium Tellurite	G12	0	-0.24516	-0.21059	-0.36968	-0.35588
Tween 40	H1	0	-1.08779	0.696663	-1.30653	-0.1099
gamma-Amino-Butyric Acid	H2	0	-1.21775	0.774571	-1.36956	-0.08862
alpha-Hydroxy-Butyric Acid	H3	0	-0.60899	0.068141	-2.00394	-0.12947
beta-Hydroxy-D,L-Butyric Acid	H4	0	-1.13932	0.812802	-1.21755	0.03327
alpha-Keto-Butryic Acid	H5	0	-1.52777	0.633552	-1.65384	-0.19834
Acetoacetic Acid	H6	0	-1.0846	0.829621	-1.27708	0.087298
Propionic Acid	H7	0	-1.21923	0.5145	-1.38129	0.072061
Acetic Acid	H8	0	-2.12981	0.217117	-2.21971	-0.13353

Formic Acid	H9	0	-0.69437	0.415616	-0.87795	0.111374
Aztreonam	H10	0	-0.55936	0.702606	-1.07294	-1.22306
Sodium Butyrate	H11	0	-1.30566	-0.4613	-1.24049	-1.32423
Sodium Bromate	H12	0	-1.04674	-0.94965	0.024105	-0.37785

Table 3. 6S Phenotypic microarray log 2 fold difference of bacterial growth between FA mutants and WT measured at 750nm

750 log2 ratio (16h)		WT	UFA-B5	CFA-B4	DUFA-B5	IBFA-B6
Negative Control	A1	0	-1.18994	0.136856	-1.11864	0.088579
Dextrin	A2	0	-0.59708	-0.02434	-1.1375	-0.04043
D-Maltose	A3	0	-2.13002	-0.47817	-2.31316	-0.38793
D-Trehalose	A4	0	-1.32375	-0.25047	-2.17114	-0.28998
D-cellobiose	A5	0	-1	0.185279	-0.33535	0.057611
Gentiobiose	A6	0	-1.02186	0.157822	-1.03014	0.061062
Sucrose	A7	0	-1.01911	0.17113	-1.1275	0.002709
D-Turanose	A8	0	-1.09466	0.131829	-1.13433	-0.08297
Stachyose	A9	0	-1.0789	0.095129	-1.1821	-0.13696
Positive Control	A10	0	-0.75924	-0.46179	-1.44996	-0.38974
pH 6	A11	0	-1.06931	-0.49192	-1.48823	-0.5454
pH 5	A12	0	-0.44224	0.140862	-0.0424	-0.65659
D-Raffinose	B1	0	-0.90388	0.263981	-1.09994	-0.04617
alpha-D-Lactose	B2	0	-0.82201	-0.23739	-2.12417	-0.26975
D-Melibiose	B3	0	-1.84212	-0.07754	-1.94301	-0.40226

Beta-methyl-D-Glucoside	B4	0	-1.26689	-0.03789	-1.33388	0.009038
D-Salicin	B5	0	-0.99594	0.144261	-1.09374	0.012141
N-Acetyl-D-Glycosamine	B6	0	-1.16077	-0.44033	-0.20972	-0.3323
N-acetyl-Beta-D-Mannosamine	B7	0	-0.91884	0.233936	-1.23025	0.084486
N-Acetyl-D-Galactosamine	B8	0	-1.14577	0.069497	-1.17932	-0.0902
N-Acetyl Neuraminic Acid	B9	0	-1.73654	-0.17242	-2.1473	-0.30183
1% NaCl	B10	0	-0.6891	-0.06885	-0.80851	0.178989
4% NaCl	B11	0	-1.98934	-0.15438	-1.0606	-0.34139
8% NaCl	B12	0	-1.52229	0.222392	-1.22023	-1.22456
alpha-D-Glucose	C1	0	-0.29913	-0.39488	-0.18649	-0.30459
D-Mannose	C2	0	-1.27954	-0.55666	-0.98929	-0.44381
D-Fructose	C3	0	-1.29869	3.24E-09	-1.73602	-0.13382
D-Galactose	C4	0	-1.84514	-0.21042	-2.22633	-0.28117
3-Methyl Glucose	C5	0	-0.6319	0.106568	-0.82294	-0.02081
D-Fucose	C6	0	-0.04295	-0.08518	0.067639	-0.04494
L-Fucose	C7	0	-1.54448	0.338824	-1.63351	-0.11938
L-Rhamnose	C8	0	-1.60363	0.157925	-1.66994	-0.08497
Inosine	C9	0	-1.5444	0.038654	-1.553	-0.2403
1% Sodium Lactate	C10	0	-1.14367	0.366389	-0.71601	-0.38417
Fusidic Acid	C11	0	-0.72918	-0.49404	-0.41704	-1.38296
D-Serine	C12	0	0.046149	0.013727	-0.09403	-0.06777
D-Sorbitol	D1	0	-2.2167	-0.37796	-2.22844	-0.4353
D-Mannitol	D2	0	-0.99793	-0.34463	-0.37185	-0.33742

D-Arabitol	D3	0	-1.13409	-0.03956	-1.00122	0.006085
myo-Inositol	D4	0	-1.09839	0.128096	-0.99599	-0.02978
Glycerol	D5	0	-1.06551	0.14056	-1.41323	0.00217
D-Glucose-6-PO4	D6	0	-0.44947	-0.1388	-0.22455	-0.46132
D-Fructose-6-PO4	D7	0	-0.77769	-0.23812	-1.3124	-0.4808
D-Aspartic Acid	D8	0	-1.24412	0.084156	-1.2042	-0.1103
D-Serine	D9	0	-2.01226	-0.22827	-2.03857	-0.31316
Troleandomycin	D10	0	-0.17602	-0.01403	-1.03371	-0.08533
Rifamycin SV	D11	0	-0.50485	-0.20148	-1.19524	-0.26219
Minocycline	D12	0	-0.1166	0.314742	-0.15599	-0.22239
Gelatin	E1	0	-1.16467	0.16145	-1.04879	0.025514
Glycyl-L-Proline	E2	0	-1.50848	-0.05162	-1.39954	-0.00398
L-Alanine	E3	0	-1.1534	0.080509	-1.20163	0.173545
L-Arginine	E4	0	-0.97481	0.230436	-0.96376	0.146204
L-Aspartic Acid	E5	0	-1.16525	0.319177	-0.99875	0.045491
L-Glutamic Acid	E6	0	-1.12229	0.11273	-1.02831	-0.07992
L-Histidine	E7	0	-0.8446	0.141438	-0.96332	0.081258
L-Pyroglutamic Acid	E8	0	-1.13929	0.128188	-1.06262	0.027818
L-Serine	E9	0	-1.5841	-0.18982	-1.73325	-0.1228
Lincomycin	E10	0	-0.1367	0.395585	-0.26497	0.105035
Guanidine HCl	E11	0	-0.19512	0.14589	-0.76503	0.004568
Niaproof 4	E12	0	-1.01778	-0.58725	-0.848	-0.44356
Pectin	F1	0	-0.80972	-0.03726	-0.86021	-0.04797

D-Galacturonic Acid	F2	0	-2.27311	-0.29678	-2.29944	-0.40911
L-Galactonic Acid Lactone	F3	0	-2.1641	-0.20167	-2.07937	-0.28619
D-Gluconinc Acid	F4	0	-1.1013	-0.55435	-1.87801	-0.32443
D-Glycyronic Acid	F5	0	-2.06619	-0.22775	-2.30762	-0.43191
Glucuronamide	F6	0	-1.20645	0.085198	-1.226	-0.13079
Mucic Acid	F7	0	-2.12151	-0.26017	-2.09465	-0.3072
Quinic Acid	F8	0	-1.1335	0.254746	-1.15174	0.04761
D-Saccharic Acid	F9	0	-1.53757	0.316817	-1.61359	-0.25794
Vancomycin	F10	0	-0.28804	0.110107	-1.29702	-1.26847
Tetrazolium Violet	F11	0	0.237781	0.723761	-0.71944	0.135367
Tetrazolium Blue	F12	0	0.007633	0.772652	-0.92609	-0.09563
p-Hydroxy-Phenylacetic Acid	G1	0	-1.30121	0.102315	-1.27368	-0.14837
Methyl Pyruvate	G2	0	-1.24151	-0.2195	-1.61445	-0.29849
D-Lactic Acid Methyl Ester	G3	0	-1.0725	0.263448	-1.09088	0.09016
L-Lactic Acid	G4	0	-1.68137	-0.37536	-1.76512	-0.27488
Citric Acid	G5	0	-1.38165	-0.0532	-1.3681	-0.13918
alpha-Keto-Glutaric Acid	G6	0	-1.29505	0.064948	-1.2391	0.14323
D-Malic Acid	G7	0	-1.07115	0.389249	-1.11149	0.097599
L-Malic Acid	G8	0	-1.52772	-0.10719	-1.2188	-0.01173
Bromo-Succinic Acid	G9	0	-1.2554	0.254192	-1.26619	0.005616
Nalidixic Acid	G10	0	-0.26182	-0.2497	-0.25696	-0.37831
Lithium Chloride	G11	0	-2.24712	-0.76499	-1.67764	-2.09383
Potassium Tellurite	G12	0	-0.31662	-0.25994	-0.3706	-0.37561

Tween 40	H1	0	-0.98852	0.148773	-1.1352	-0.10887
gamma-Amino-Butyric Acid	H2	0	-1.12857	0.090893	-1.22761	-0.10932
alpha-Hydroxy-Butyric Acid	H3	0	-0.95398	-0.15431	-1.48543	-0.23865
beta-Hydroxy-D,L-Butyric Acid	H4	0	-1.05037	0.17381	-1.09867	0.055655
alpha-Keto-Butryic Acid	H5	0	-0.78695	0.550543	-0.89876	0.127419
Acetoacetic Acid	H6	0	-0.80474	0.041223	-0.71965	-0.18015
Propionic Acid	H7	0	-0.80985	0.341496	-0.89772	0.189056
Acetic Acid	H8	0	-1.05153	0.200533	-1.08287	0.274684
Formic Acid	H9	0	-0.67897	0.241818	-0.7761	0.141737
Aztreonam	H10	0	-0.09282	-0.07192	-0.15737	-0.15249
Sodium Butyrate	H11	0	-1.39793	-0.56814	-1.12453	-1.28847
Sodium Bromate	H12	0	-1.46256	-1.25908	-0.18793	-0.67236

Table 3. 7S Comparison of unsaturated fatty acid and IBFA production strategies across multiple works. ND: Not Determined. E: Heterologous expression, O: Overexpress; KO: Knockout, CRISPRi: CRISPR interference/knockdown

Target: ω7-unsaturated fatty acids					
Strain Name	Genetic Modifications	*E coli* strain Background	ω7-UFA Titer (mg L^{-1})	ω7-UFA %	Source
fabAB+TE	O: *fabA, fabB, AtFatA*	BL21(DE3)	23.6	47.9	Cao, 2010(33)
FRT-ΔfadE	O: *fadR, tesA* KO: *fadE*	DH1	2236	43	Zhang, 2012(40)
UFA1	O: *fabB*	BW27783 2T	ND	~70	Budin, 2018(85)

fabB93	O: *fabB*	MG1655	ND	12	Santoscoy, 2021(74)
ω7-UFA-3	O: *fabB, fadR* KO:*fadE, cfa*	MG1655	131.8	75.8	This work
ω7-UFA-5	O: *fadR* KO:*fadE, fabR, cfa*	MG1655	197.8	69.5	This work

Target: cyclopropane fatty acids[a]					
Strain Name	Genetic Modifications	*E coli* strain Background	CFA Titer ($mg\ L^{-1}$)	CFA %	Source
MG1655 (WT)	None	MG1655	ND	38.6	Yao Chen, 2016(86)
MG1655 (WT)	None	MG1655	21.0	12.4	This Work
CFA-4	O: *fadR, cfa_{Ec}* *E: cfa* (C. *butyricum)* KO:*fadE, fabR*	MG1655	84.7	55.3	This Work

Target: Δ5-monounaturated fatty acids					
Strain Name	Genetic Modifications	*E coli* strain Background	Δ5-MUFA Titer ($mg\ L^{-1}$)	Δ5-MUFA %	Source
pET22b-desA	E: des (B. subtilis)	BL21(DE3)	N.D.	21.9	Bonamore, 2006(47)
pBR-fabA-des	O: *fabA* E: des *(B. subtilis*)	W3110	N.D.	7.7	Luo, 2009(87)
des93	E: *des (B. subtillus)*	MG1655	N.D.	20.3[b]	Santoscoy, 2021(74)
Δ5-UFA-3	O: *fadR*	MG1655	75.5	59.8	This Work

	KO: *fadE, cfa* E: des (B. subtilis), Fd_{Ec} CRISPRi: *fabB*				
Target: Double unsaturated fatty acids					
Strain Name	Genetic Modifications	*E coli* strain Background	DUFA Titer ($mg\ L^{-1}$)	DUFA %	Source
pET22b-desA	E: des (B. subtilis)	BL21(DE3)	N.D.	12.9	Bonamore, 2006(47)
DUFA-5	O: *fadR* KO: *fadE, cfa, fabR* E: *des (B. subtilis)*	MG1655	158.4	46.5	This Work
Target: Internally branched FAs					
Strain Name	Genetic Modifications	*E coli* strain Background	IBFA Titer FA ($mg\ L^{-1}$)	IBFA %	
BfaAB + exogenous 18:1 Δ9	E: *BfaA BfaB* (M. tuberculosis)	Rosetta 2	ND	0.7	Machida, 2017(44)
IBFA-5	O: fadR KO: fadE, cfa, fabR E: BfaAB (*M. smegmatis*)	MG1655	91.3	39.2	This work

[a] No comparable works for *E. coli* CFA overproduction

[b] Total UFA, Δ5-MUFAs and ω7-UFAs not distinguished

3.6.2 Figures

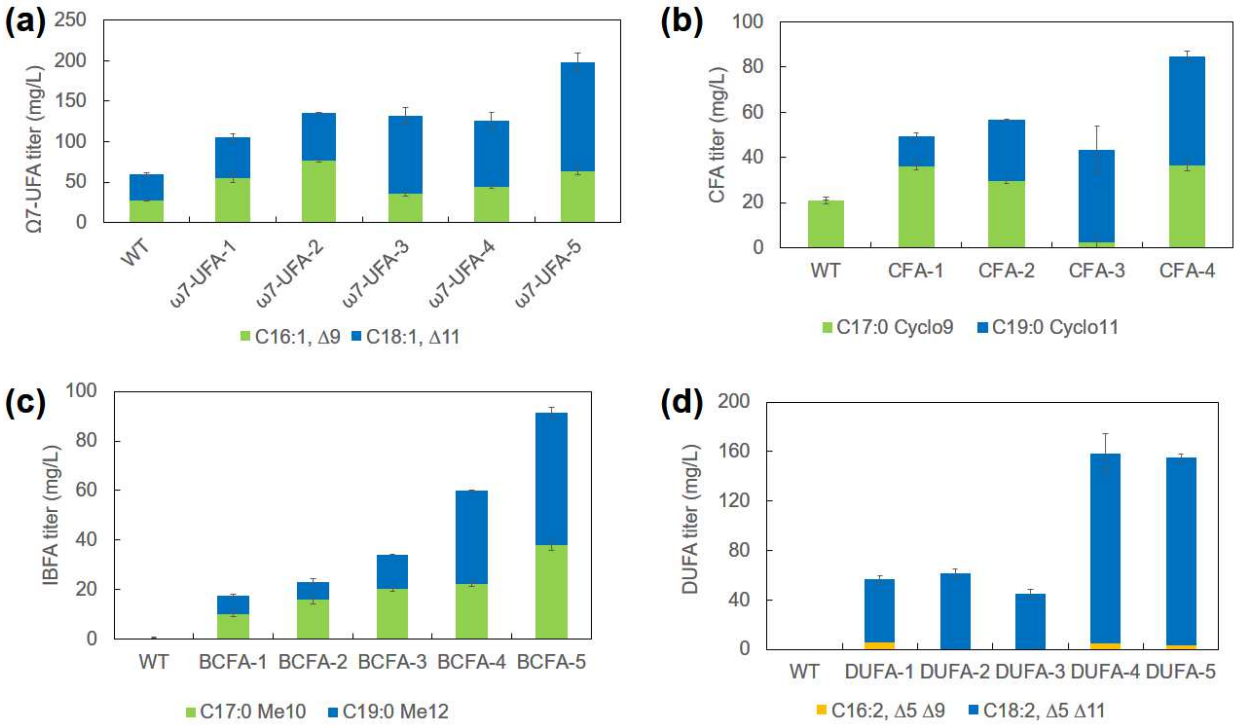


Figure 3. 10S FA titers for novel strains. 1 Titers of ω7-UFA (a), CFA (b), IBFA (c), and DUFA (d) in engineering E. coli strains. Error bars represent standard deviation measured from biological triplicates.

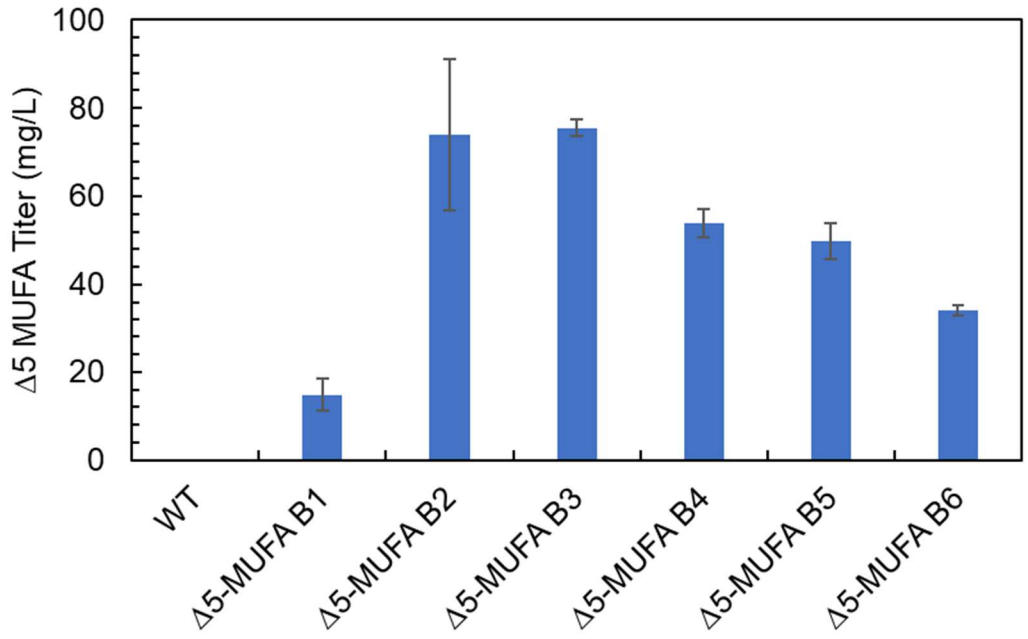


Figure 3. 11S Titer of Δ5-MUFA in engineered strains. Error bars represent standard deviation measured from biological triplicates.

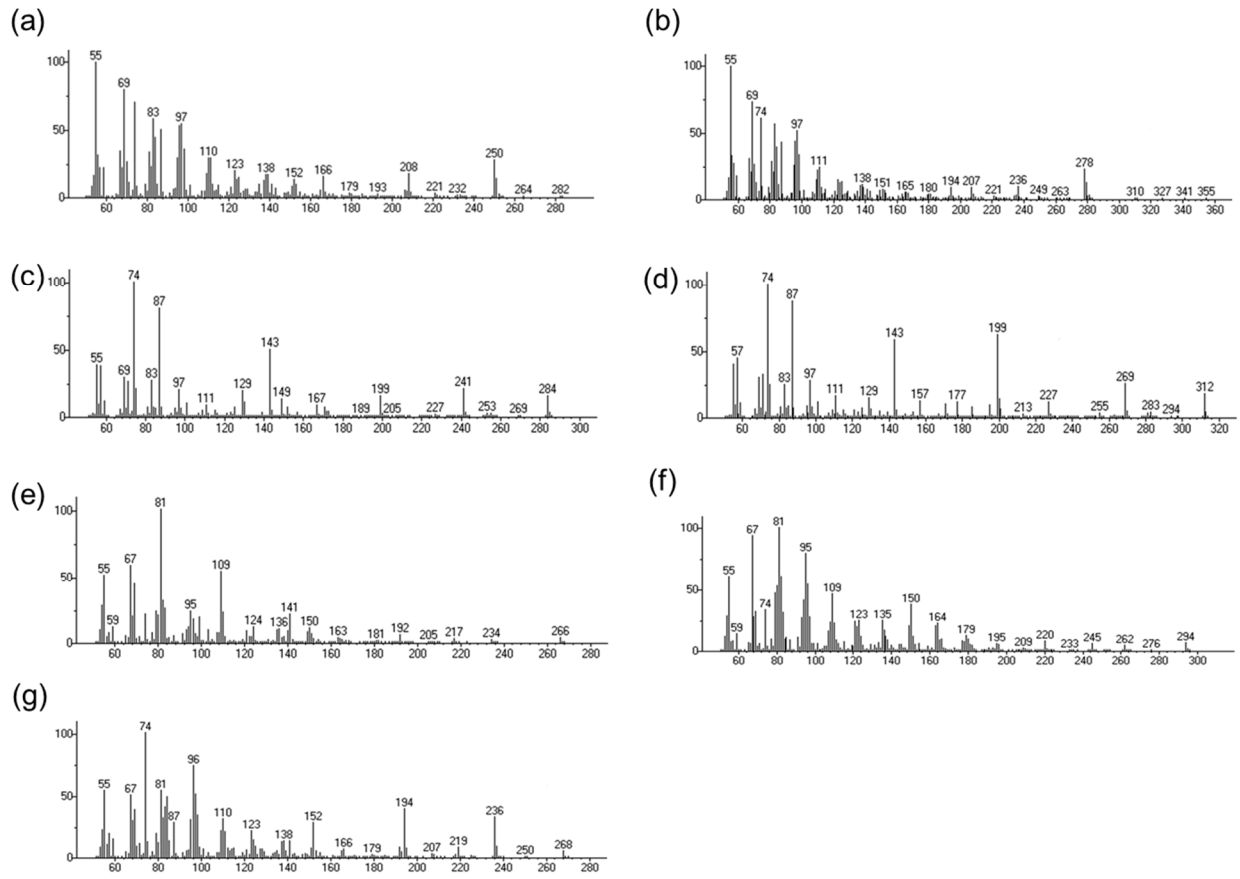


Figure 3. 12S The mass spectra of the modified fatty acids. a C17:0 CFA. b. C19:0 CFA. c. C17:0Me10. d. C19:0Me12. e. C16:2 Δ5, 9. f. C18:2, Δ5Δ11. g. C16:1 Δ5.

GEN III MicroPlate

A1 Negative Control	A2 Dextrin	A3 D-Maltose	A4 D-Trehalose	A5 D-Cellobiose	A6 Gentiobiose	A7 Sucrose	A8 D-Turanose	A9 Stachyose	A10 Positive Control	A11 pH 6	A12 pH 5
B1 D-Raffinose	B2 α-D-Lactose	B3 D-Melibiose	B4 β-Methyl-D-Glucoside	B5 D-Salicin	B6 N-Acetyl-D-Glucosamine	B7 N-Acetyl-β-D-Mannosamine	B8 N-Acetyl-D-Galactosamine	B9 N-Acetyl Neuraminic Acid	B10 1% NaCl	B11 4% NaCl	B12 8% NaCl
C1 α-D-Glucose	C2 D-Mannose	C3 D-Fructose	C4 D-Galactose	C5 3-Methyl Glucose	C6 D-Fucose	C7 L-Fucose	C8 L-Rhamnose	C9 Inosine	C10 1% Sodium Lactate	C11 Fusidic Acid	C12 D-Serine
D1 D-Sorbitol	D2 D-Mannitol	D3 D-Arabitol	D4 myo-Inositol	D5 Glycerol	D6 D-Glucose-6-PO4	D7 D-Fructose-6-PO4	D8 D-Aspartic Acid	D9 D-Serine	D10 Troleandomycin	D11 Rifamycin SV	D12 Minocycline
E1 Gelatin	E2 Glycyl-L-Proline	E3 L-Alanine	E4 L-Arginine	E5 L-Aspartic Acid	E6 L-Glutamic Acid	E7 L-Histidine	E8 L-Pyroglutamic Acid	E9 L-Serine	E10 Lincomycin	E11 Guanidine HCl	E12 Niaproof 4
F1 Pectin	F2 D-Galacturonic Acid	F3 L-Galactonic Acid Lactone	F4 D-Gluconic Acid	F5 D-Glucuronic Acid	F6 Glucuronamide	F7 Mucic Acid	F8 Quinic Acid	F9 D-Saccharic Acid	F10 Vancomycin	F11 Tetrazolium Violet	F12 Tetrazolium Blue
G1 p-Hydroxy-Phenylacetic Acid	G2 Methyl Pyruvate	G3 D-Lactic Acid Methyl Ester	G4 L-Lactic Acid	G5 Citric Acid	G6 α-Keto-Glutaric Acid	G7 D-Malic Acid	G8 L-Malic Acid	G9 Bromo-Succinic Acid	G10 Nalidixic Acid	G11 Lithium Chloride	G12 Potassium Tellurite
H1 Tween 40	H2 γ-Amino-Butryric Acid	H3 α-Hydroxy-Butyric Acid	H4 β-Hydroxy-D,L-Butyric Acid	H5 α-Keto-Butyric Acid	H6 Acetoacetic Acid	H7 Propionic Acid	H8 Acetic Acid	H9 Formic Acid	H10 Aztreonam	H11 Sodium Butyrate	H12 Sodium Bromate

Figure 3. 13S Culture media and conditions of the GEN III MicroPlate used for phenotypic profiling of engineered E. coli strains.

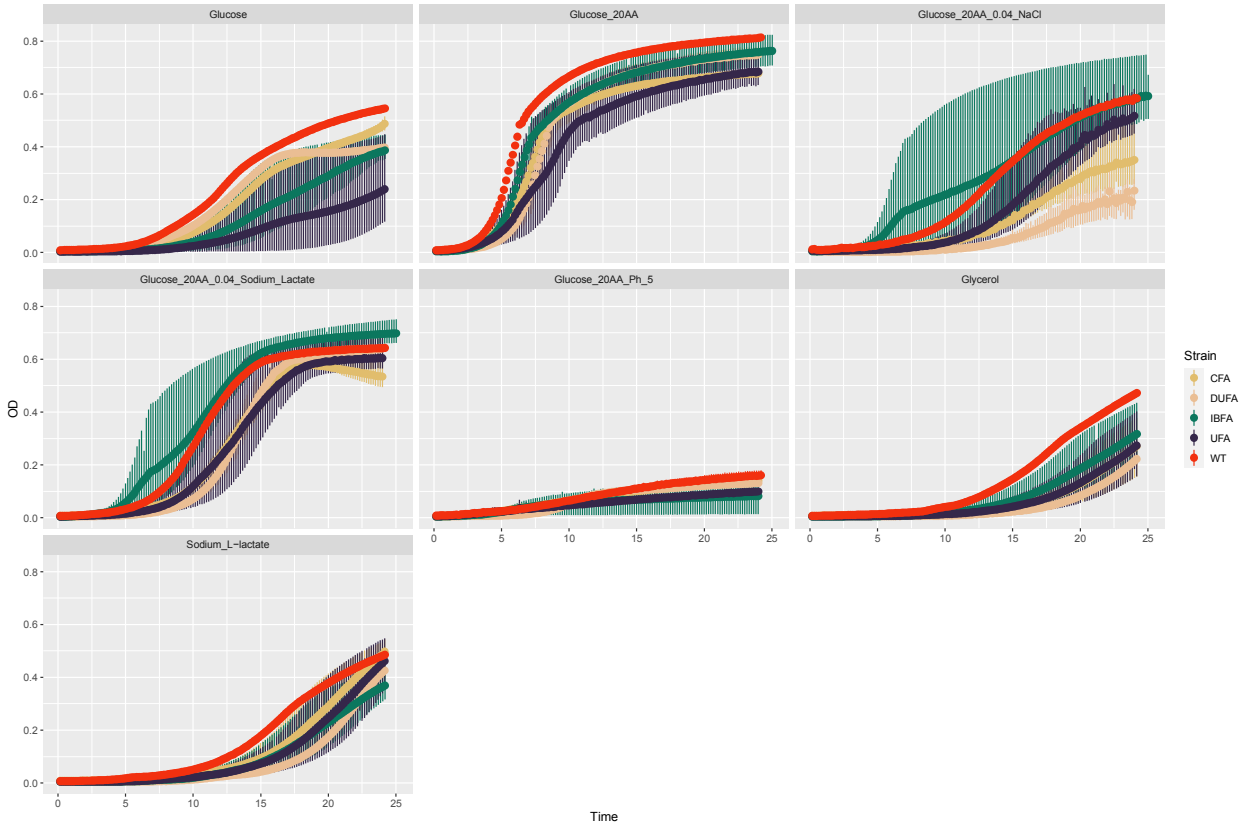


Figure 3. 14S Growth curves of each novel strain compared to WT in different environmental conditions. Horizontal colored lines are +/- standard deviation bars for each time point. Lines represent the mean of triplicates for each strain in each condition.

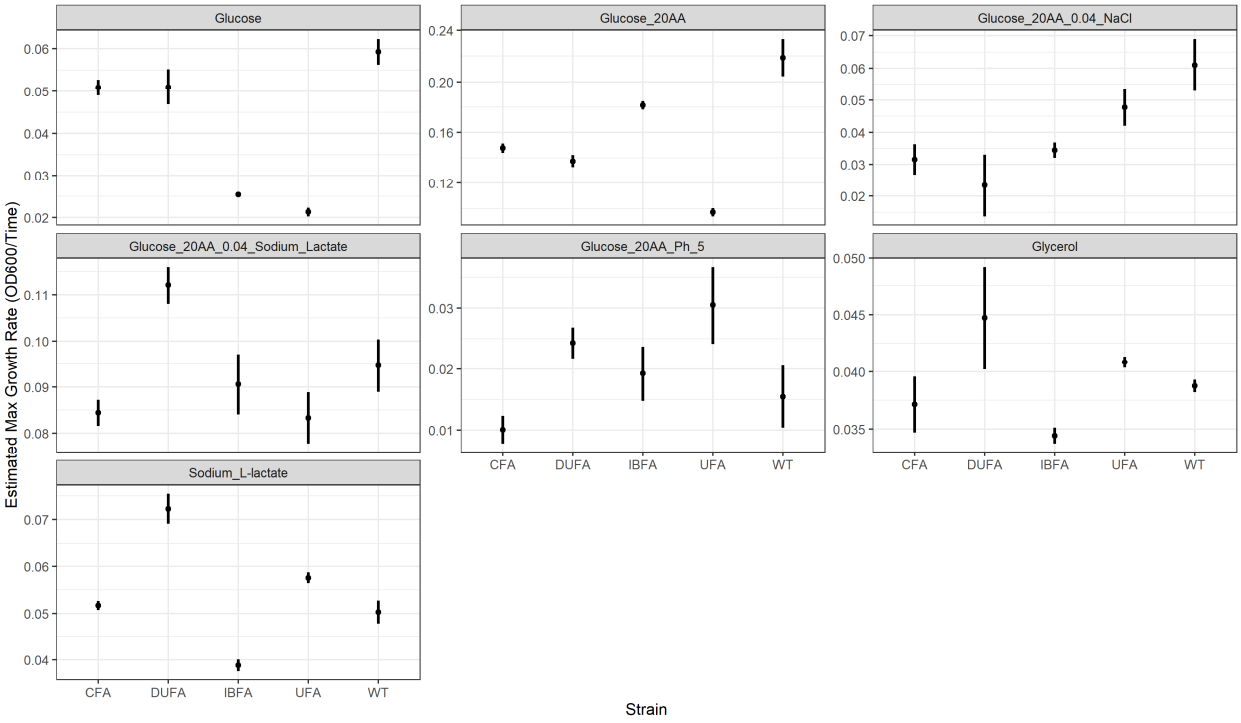


Figure 3. 15S Estimated maximal growth rate of each novel strain compared to WT in different environmental conditions. A sliding window was used to calculate slope for every 15-minute interval during the experiment, and a loess regression was fit to the first derivative, and the maximum value of this regression found. Lines signify 95% confidence intervals.

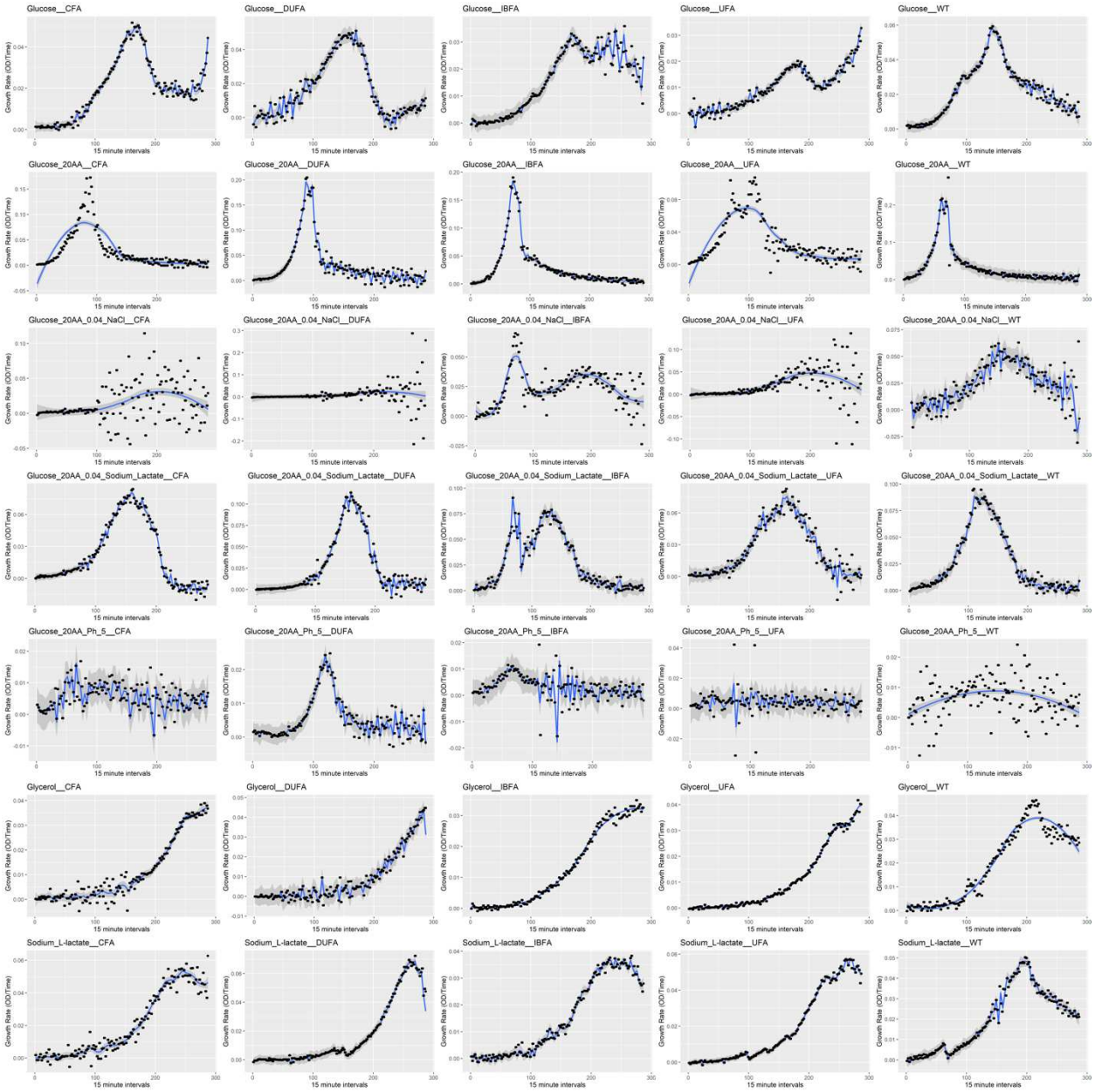


Figure 3. 16S First derivative of bacterial growth (OD600) calculated at 15-minute intervals. A loess regression was fit to this data, and 95% confidence intervals calculated and shaded in grey.

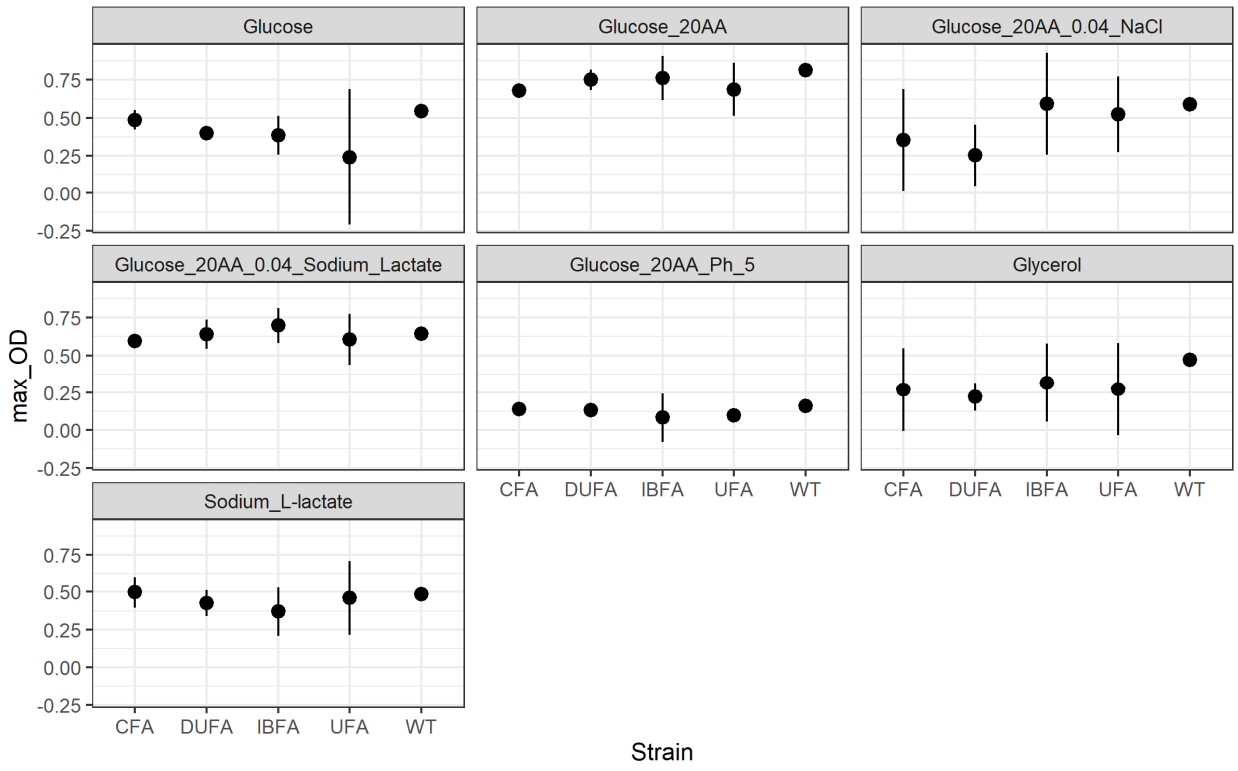


Figure 3. 17S Maximal cell density (OD600) of each novel strain compared to WT in different environmental conditions. Black dots visualize the mean with black lines signifying 95% confidence intervals estimated from triplicates.

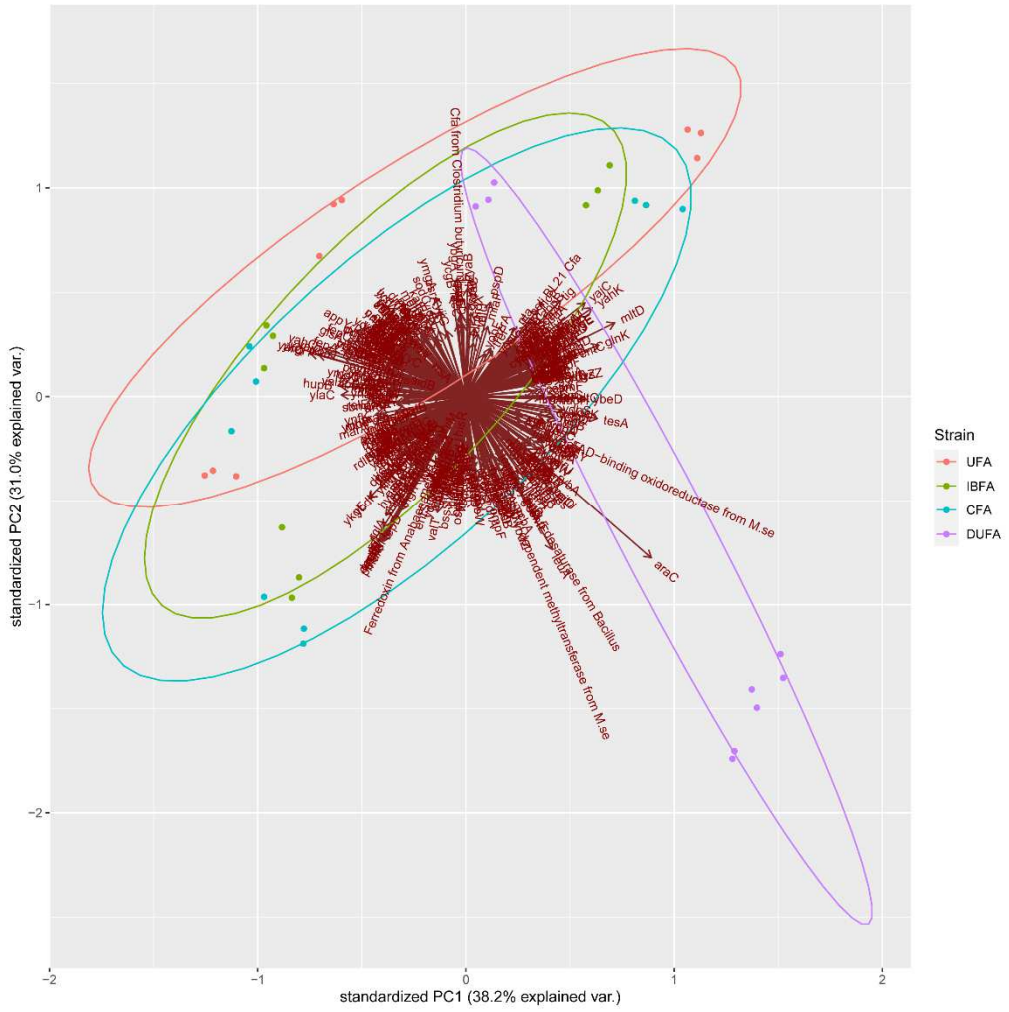


Figure 3. 18S Biplot of a principal component analysis of transcriptomic samples from all strains. Red text identify transcript loading variable direction and magnitude.

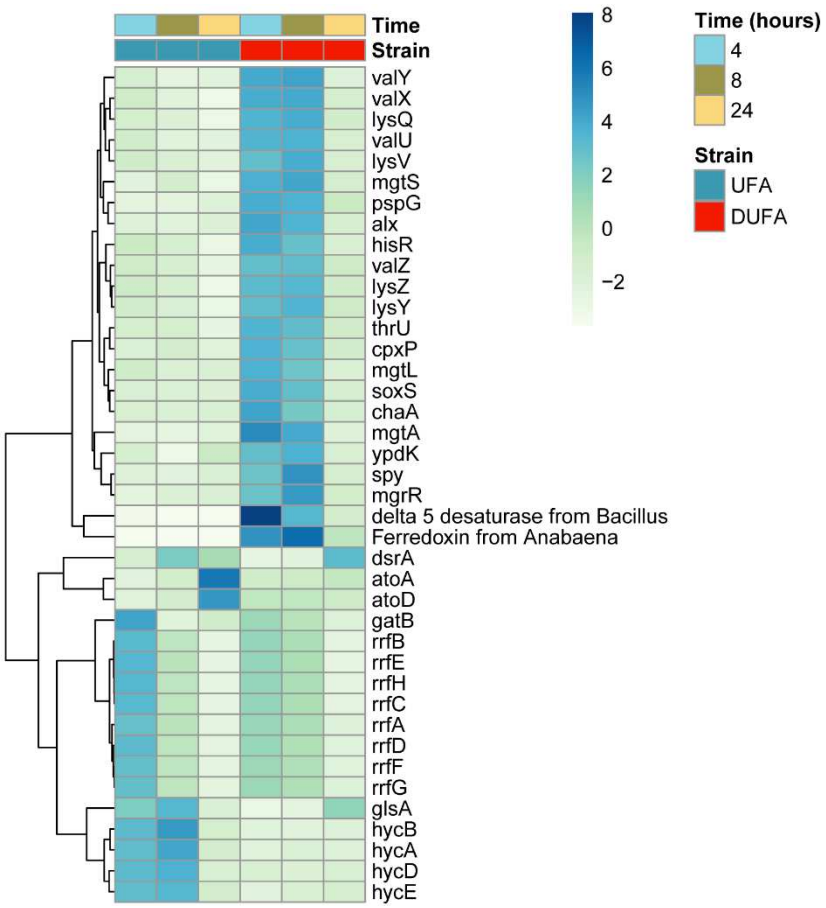


Figure 3. 19S Highly variable transcripts across time when comparing DUFA to WT expression. Expression values from DESeq2 analysis are clustered using the minkowski metric.

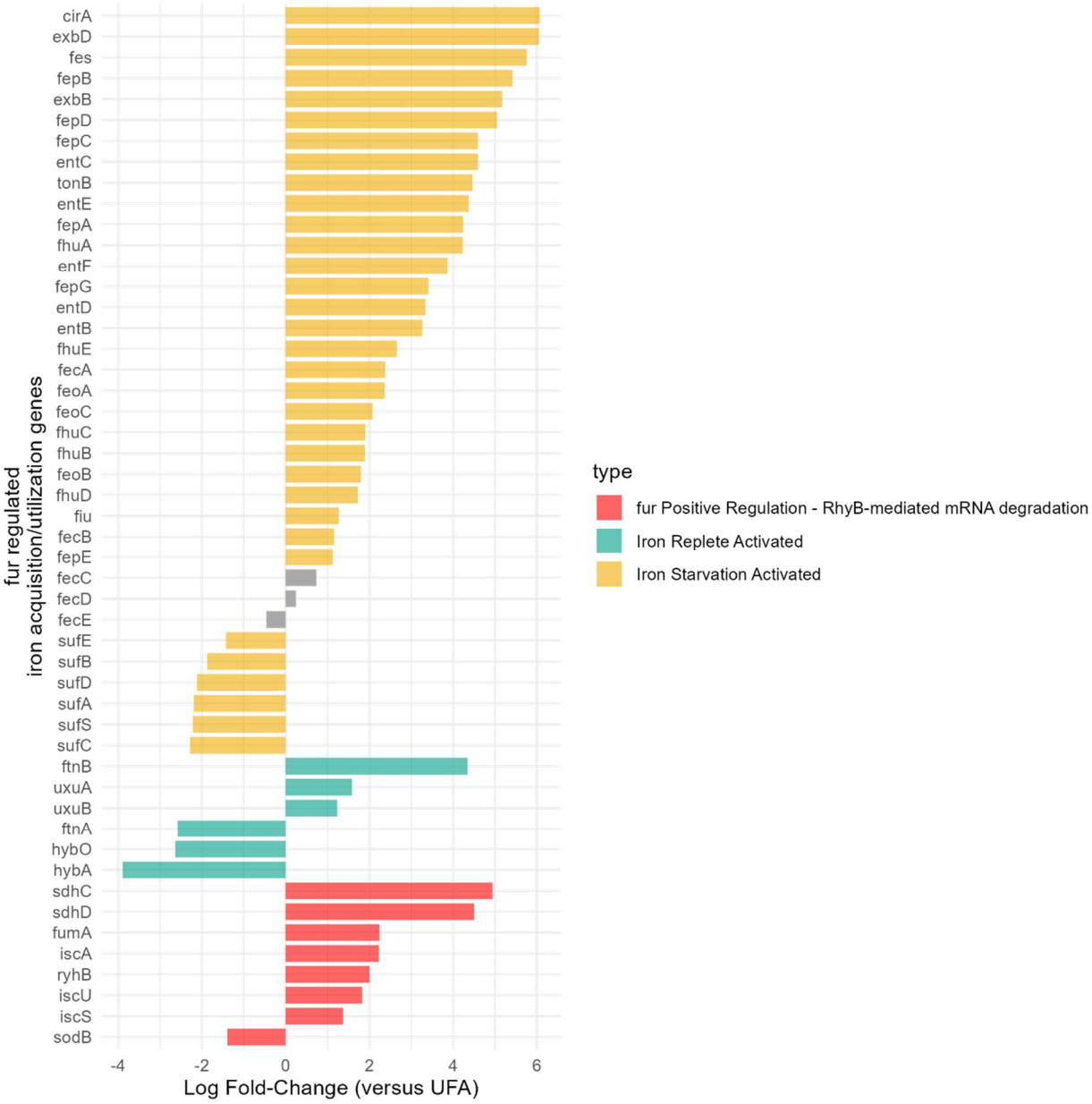


Figure 3. 20S Fur regulated iron intake and acquisition genes are differentially expressed in DUFA at 4 hours. Log-fold changes for each gene compared to UFA expression. Colored groupings represent the type of fur regulation or conditions for expression. rhyB is not positively regulated by fur, but is grouped within the fur positive regulation regulon as these genes are also degraded via rhyB and their expression is negatively correlated with rhyB expression. Grey bars indicate non-significant fold changes. All genes included were experimentally validated as targets in Seo et al. 2014.(79)

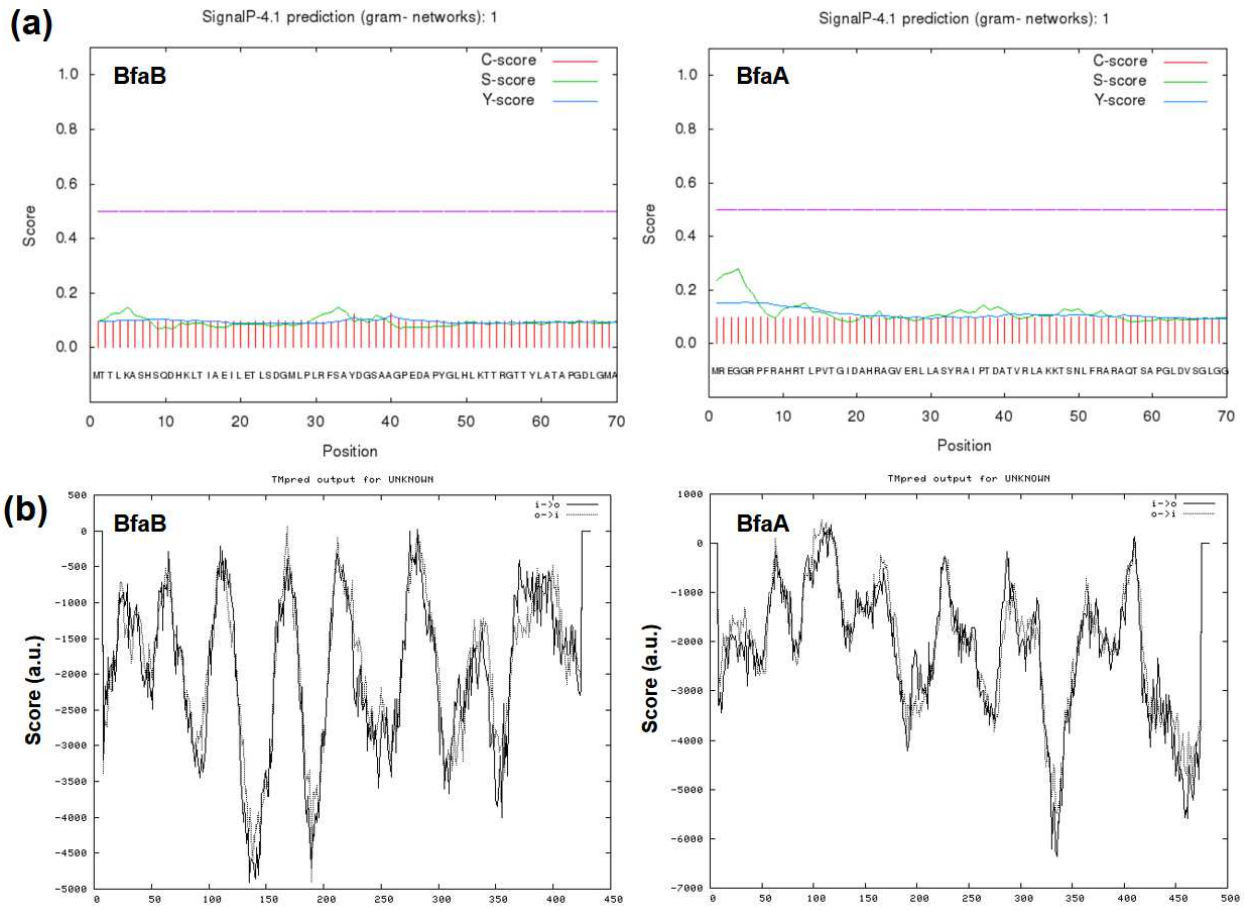


Figure 3. 21S Prediction of possible secretion tags and transmembrane sequences for BfaB and BfaA enzymes from Rhodococcus opacus. a. Secretion tag prediction was performed using the SignalP 4.1 Server(88). A protein with all score values (C-, S-, and Y-scores) close to 0.1 is regarded as a non-secretory protein. b, Prediction of transmembrane sequences was performed on the TMpred server(89). Only scores above 500 are considered significant for possible transmembrane sequence.

Chapter 4:

Autologous transcription regulator overexpression for increased triacylglyceride production in *Rhodococcus opacus PD630*

4.1 Introduction

4.1.1 Abstract

Lignocellulosic biomass is currently underutilized but offers promise as a resource for generation of commercial end-products, such as biofuels, detergents, and other oleochemicals. *Rhodococcus opacus PD630* is an oleaginous, gram-positive bacterium with an exceptional ability to degrade recalcitrant lignin breakdown products to produce triacylglycerides, an important biofuel precursor. Triacylglycerides are formed as carbon storage coupled to environmental nitrogen levels, resulting in accumulation only during times of low growth. In this work, we implement a top-down genetic screen to identify native transcriptional regulators upregulated in low nitrogen. We successfully creating three novel strains of *Rhodococcus opacus PD630* with increased lipid production in nitrogen replete conditions, even when grown on phenol, an aromatic lignin breakdown product. Transcriptomic analysis reveals signatures of carbon source specific and universal DE for increasing lipid production, tied to repression of the phenylacetic acid degradation pathway. Gene deletion experiments confirm the existence of a complex regulatory mechanism tying multiple branches of phenylalanine metabolism to lipid accumulation. By generating mutants uncoupling carbon storage from nitrogen concentration, we

move closer towards optimizing *R. opacus PD630* for efficient bioproduction on lignocellulosic biomass.

4.1.2 Introduction:

Utilization of biomass, especially lignocellulosic biomass, has enormous unmet potential – biofuels and commodity chemicals are commonly produced from the cellulose and hemicellulose fractions, but the lignin fraction is most frequently treated as a waste stream and burned(1). Lignin offers promise as a feedstock for microbial bioproduction of biofuels and platform chemicals; it is both energy-dense and has a high carbon-to-oxygen ratio (greater than 2:1) (2, 3). Yet achieving commercially viable bioproduction is not without challenges as lignin is a complex, heterogeneous polymer, and the depolymerization process primarily yields aromatic compounds toxic to many microbes(4, 5).

Unlike most prokaryotes, Rhodococcus species have the capacity to produce triacylglycerols (TAGs) as a storage compound, rather than the more common polyhydroxyalkanoates (PHAs)(6-8). *R. opacus PD630* (hereafter *R. opacus*) can accumulate more than half of its dry cell weight (dcw) as TAGs when grown on gluconate(8), and as much as 44% of dcw when grown on phenol(9), a monoaromatic commonly found in lignin. TAGs and their derivatives comprise a substantial portion of industrial chemicals, representing an area of bioproduction where finely-tuned microbial cell factories could make a significant impact(10).

Lipid production in oleaginous Rhodococcus species is largely linked to nitrogen stress: in nitrogen poor growth conditions, cells store carbon in the form of TAGs(11, 12). When the nutrient situation is reversed and carbon becomes scarce in the immediate environment, these storage molecules can then be readily mobilized. On a systemic level, linking lipid production to

nitrogen starvation necessarily decouples it from growth, with most TAG accumulation corresponding to stationary phase(13). This nitrogen paucity presents a challenge for bioproduction on recalcitrant biomass; lipid titers on aromatic substrates can be an order of magnitude lower than on glucose due to additive stress impairing growth(13). In *Rhodococcus*, most strategies to improve TAG production have focused on overexpressing native genes, including: 1. increasing fatty acid synthesis with the fasI operon, 2. boosting the final step in TAG biosynthesis with atf2, 3. using thioesterases to increase fatty acid-CoA production, and 4. increasing the NAD(P)H pools via tadD or autologous malic enzymes(11, 14-17). While these methods have yielded lipid accumulation in oleaginous rhodococci, none of them address the nitrogen-dependence of peak TAG production, nor attempt to cultivate a production strain on a toxic feedstock like lignin or its byproducts.

Here, top-down genomics are used to identify three regulatory targets predicted to increase lipid titers in a nitrogen-independent manner. Transcriptomic analysis of mutant strains grown in a nitrogen rich environment with phenol as a sole carbon source identified two novel expression states during increased lipid production, with significantly increasing TAG production in phenol through increased expression of phenolic utilization genes. Fermentation in permissive versus limiting carbon and nitrogen conditions confirmed that overexpressing the phenylacetic acid degradation regulator paaX promoted fatty acid accumulation in a nitrogen-independent manner. Molecular interrogation of this phenotype revealed a complex regulatory mechanism whereby repression of the phenylacetate degradation pathway requires expression of the feaR activator of the upstream phenylethylamine degradation pathway for lipid accumulation in nitrogen rich environments. In conclusion, we identified targets for enhancing *R. opacus* for

high lipid titer in nitrogen-replete conditions and successfully optimized *R. opacus* as a chassis for bioproduction from recalcitrant feedstocks.

4.3 Methods:

4.3.1 Chemicals & Strains:

Unless otherwise indicated, all chemicals were purchased from Sigma Aldrich (St. Louis, MO). The ancestral, or wild-type (WT), strain for all transformant cells lines was *Rhodococcus opacus* PD630 (DSMZ 44193); this strain was used as a basis of comparison for all transgenic strains. Culturing conditions for all experiments, unless otherwise noted, were incubation at 30°C with 250rpm shaking, with the previously-described minimal salts medium B constituting the growth medium[50]. Media was sterilized using a 0.22µm filter, with carbon sources added as filter-sterilized stock solutions; nitrogen was added either pre-sterilization or as a separate filter-sterilized stock solution. Media pH was adjusted to 7.2 using 6N HCl or 2M NaOH solutions. Optical density at 600nm (OD_{600}) was measured using a Tecan Infinite 200Pro plate reader, either directly using VWR semi-micro polystyrene cuvettes or indirectly based on the absorbance at 600nm (A_{600}) measured in black 96-well plates (Greiner Bio-One flat bottom, chimney well, µclear); an A_{600} value can be converted into an OD_{600} value (for *R. opacus* cultures) via the experimentally-determined relationship $OD_{600} = 1.975 \times (A_{600} - 0.04)$.

All strains were maintained on tryptic soy broth (TSB) plates supplemented with 1.5% agar. Kanamycin (20µg/mL), gentamicin (10µg/mL), chloramphenicol (34µg/mL), or hygromycin B (200µg/mL) was added as appropriate to *E. coli* cultures. Kanamycin (50µg/mL), gentamicin (10µg/mL), chloramphenicol (15µg/mL), and/or hygromycin B (50µg/mL) were added as appropriate to *R. opacus* cultures.

4.3.1 Plasmid Construction and DNA Manipulation:

All plasmids constructed for this study were confirmed by DNA sequencing (Genewiz; South Plainfield, NJ); all primers were purchased from Integrated DNA Technologies (IDT; Coralville, IA). All overexpression plasmids were assembled using GoldenGate Assembly, and knockout plasmids were assembled using Gibson Assembly; all plasmids were replicated in *E. coli* DH10B, then isolated using a PureLink™ HiPure Plasmid Miniprep Kit (Invitrogen by ThermoFisher; Waltham, MA)[51,52]. DNA fragments amplified were amplified using Phusion High-Fidelity DNA Polymerase (NEB; Ipswich, MA) and purified using a ZymoClean Gel DNA Recovery Kit (Irvine, CA); genomic DNA was extracted from *R. opacus* using a Promega Wizard™ Genomic DNA Purification Kit (Madison, WI).

4.3.2 Transformation of *R. opacus*:

Preparation of competent cells was conducted as previously described[50]. In brief, an overnight culture in TSB medium was used to inoculate 100mL of fresh TSB containing 8.5g/L glycine and 10g/L sucrose (initial optical density, OD_{600}, diluted to 0.075); cells were cultivated in standard conditions to an OD_{600} value of 0.4-0.5, corresponding to exponential phase growth. Cells were rapidly chilled and centrifuged at 3.5k relative centrifugal force (rcf), washing twice with chilled, sterile, deionized water; a final resuspension to an OD_{600} ~ 10-15 was conducted in chilled 10% (v/v) glycerol, and cells were aliquoted at 100µL and frozen at -80°C for transformation.

For transformation with a replicating plasmid, approximately 500ng plasmid DNA was added to prepared electrocompetent cells; cells were shocked at 2500mV across a 0.2cm-gap cuvette (time constant ~5-6ms) and washed with 1mL rich media (either TSB or SOC/super-optimal broth

with catabolites) to recover. For outgrowth, cells were transferred to 50mL glass culture tubes and incubated under standard growth conditions for 4 hours before being spread on TSB plates infused with the appropriate antibiotics. Plated cells were grown at 30°C for 2-3 days, until colonies emerged, then propagated on fresh plates.

Generation of *R. opacus* knockout mutants was accomplished using a previously-developed method for homologous recombination, with modifications[36]. Briefly, electrocompetent cells were prepared as above using a strain expressing a helper plasmid containing the Che9c viral recombinases. Competent cell aliquots were transformed with 1-2µg of a suicide vector containing the knockout construct (an antibiotic resistance cassette flanked by ~500bp segments homologous to the target gene); the outgrowth period for these transformations was at least 6 hours and up to 12 hours. Transformed cells were plated on TSB with the corresponding antibiotics and incubated at 30°C for 4-5 days; colonies were propagated on fresh TSB plates and verified by colony PCR (Promega GoTaq® G2 DNA Polymerase).

4.3.3 Fermentation for Lipid Analysis:

Frozen stocks of strains (generated from isolated colonies and stored at -80°C) were streaked onto fresh TSB plates, with antibiotics as appropriate, then grown for 2-4 days; a loopful of cells were used to inoculate seed cultures in minimal media B with 1g/L each $(NH_4)_2SO_4$ and glucose as nitrogen and carbon sources, respectively[50]. These cultures were centrifuged at 3.5k rcf and the pellets resuspended in low-nitrogen minimal media; OD_{600} of these cell suspensions was adjusted to approximately 2, then used to inoculate the 50mL fermentation cultures (250mL non-baffled Erlenmeyer flasks). The carbon conditions were either 2g/L glucose or 0.4g/L phenol, and the nitrogen conditions were either 0.05g/L or 1g/L $(NH_4)_2SO_4$ (hereafter 'low' and 'high' nitrogen), for a total of four combinatorial conditions; each strain was grown in one flask per

carbon/nitrogen condition. Cultures were grown in standard conditions for 72 hours unless otherwise described; final OD_{600} was measured and used to calculate the volume necessary to collect 5 OD units of cells (or 10 OD units, in the case of the glucose/low nitrogen condition) using the relation $Vol = 5/OD_{600}$. Triplicate samples of each strain and culture condition were collected and centrifuged at 3.5k rcf for 10 minutes, then the culture supernatant was discarded, and the cell pellets were stored at -20°C prior to lipid extraction. The low nitrogen concentration (0.05g/L $(NH_4)_2SO_4$) was chosen to optimize fatty acid titer in phenol cultures, and the high-nitrogen condition (1.0g/L $(NH_4)_2SO_4$) is sufficient for nitrogen-replete growth and fatty acid accumulation in both carbon conditions; optimal FA accumulation in glucose was not required.

4.3.4 Lipid Extraction & Analysis:

An acid-chloroform lipid extraction was performed as described in Amara et al (2016), with modifications. In brief, the pelleted 5 (or 10) OD units of cells were resuspended in 100µL sterile, deionized H_2O and transferred to a 15mL glass centrifuge tube. A 1mL aliquot of 10% (v/v) H_2SO_4 in methanol was added to the cells, as well as 1mL of chloroform and a C_{12} standard (40mg/mL lauric acid dissolved in methanol) to a final concentration of 40mg/L. Cell solutions were incubated at 100°C for one hour, then chilled rapidly on ice prior to adding 1mL of deionized H_2O and mixing thoroughly by vortex. Finally, the cell extracts were centrifuged at 1k rpm for 5 minutes at room temperature, and the resulting organic layer was extracted into GC vials.[49] Samples were stored at 4°C prior to analysis.

Lipid extracts were analyzed via GC-FID using an Agilent 6890N Network GC system with Enhanced Agilent MSD ChemStation software and equipped with a 7683B series injector, G2614A autosampler, and DB-5ms column. A 1µL sample of each lipid extract was splitlessly injected into the 250°C inlet, with N_2 carrier gas flowing at 1.4mL/min. Each sample run began

with an initial oven temperature of 80°C, then ramped at 20°C/min to 300°C, where it held for a final three minutes. Between samples, the injector was washed with ethyl acetate, with ethyl acetate also serving as the blank samples at the start and end of each sample run. Peak integration was carried out using the ChemStation software and exported to Microsoft Excel for data processing.

In addition to the C_{12} internal standard, lipid samples were compared to standard curves for thirteen other fatty acids, detailed in **Error! Reference source not found.**. All standard curves lotted the area of the peaks of 25mg/L, 50mg/L, and 100mg/L standard compounds. The fitted slopes of these curves were used to calculate the concentration of each component within the extracted lipid samples, matched by retention time and normalized to the 40mg/L lauric acid internal standard. If necessary for comparison, summed average titers were transformed to standardize control samples between different fermentation experiments.

Table 1. Fatty acid standards for analysis of lipid samples extracted from Rhodococcus opacus PD60 cells.

Short Notation	Long Notation	Linear Formula	Retention Time (min)
C12:0	lauric/dodecanoic acid	$CH_3(CH_2)_{10}COOH$	6.23
C14:0	myristic/tetradecanoic acid	$CH_3(CH_2)_{12}COOH$	7.388
C15:0	pentadecanoic acid	$CH_3(CH_2)_{13}COOH$	7.929
C16:1	palmitoleic/hexadecenoic acid	$CH_3(CH_2)_5CH=CH(CH_2)_7COOH$	8.351
C16:0	palmitic/hexadecanoic acid	$CH_3(CH_2)_{14}COOH$	8.443
C16+Me	methyl palmitate/ hexadecanoate	$CH_3(CH_2)_{14}CO_2CH_3$	8.624
C17:1	heptadecenoic acid	$CH_3(CH_2)_5CH=CH(CH_2)_8COOH$	8.804

C17:0	margaric/heptadecanoic acid	$CH_3(CH_2)_{15}COOH$	8.938
C17+Me	methyl margarate/ heptadecanoate	$CH_3(CH_2)_{15}CO_2CH_3$	9.12
C18:1	oleic/octadecenoic acid	$CH_3(CH_2)_7CH{=}CH(CH_2)_7COOH$	9.302
C18:0	stearic/octadecanoic acid	$CH_3(CH_2)_{16}COOH$	9.408
C18+Me	methyl stearate/octadecanoate	$CH_3(CH_2)_{16}CO_2CH_3$	9.57
C19:1	nonadecenoic acid	$CH_3(CH_2)_7CH{=}CH(CH_2)_8COOH$	9.74
C19:0	nonadecanoic acid	$CH_3(CH_2)_{17}COOH$	9.862

4.3.5 RNA extraction and rRNA depletion:

RNA extraction proceeded as previously described (cite Henson et al. paper). Briefly, RNA was extracted using the RNA MiniPrep kit (Zymo Research) and treated with two doses of TURBO DNase I (Ambion) for 30 minutes at 37C to remove DNA contamination. DNase-treated RNA was then cleaned using the RNA Clean & Concentrator Kit (Zymo Research), and then tested for DNA contamination by PCR amplification using intergenic primers. Any samples with distinct bands after PCR were digested and cleaned again until no DNA was detected. Total RNA concentration was quantified using a NanoVue Plus spectrophotometer and rRNA was depleted using the Bacterial Ribo-Zero rRNA removal kit (Illumina). mRNA was converted to cDNA and barcoded using previously described methods (CITE YONEDA ET AL). cDNA samples were then pooled in equimolar ratios, diluted in nuclease-free water to a final concentration of 10nM for sequencing.

4.3.6 Sequencing Library Preparation and Transcriptomic Analysis:

A 20 uL equimolar mix of cDNA samples were submitted for sequencing at the Center for Genome Sciences and Systems Biology in Washington University in St. Louis School of Medicine. Samples were single-end sequenced (1x75bp) using the Illumina Hi-Seq 2500 System.

After demultiplexing, raw reads were trimmed using trimmomatic using the standard settings and the CROP length of 75bp. Samples with more than 15 million trimmed reads were subsampled using seqtk, and then mapped to a bowtie2 library built off the ASM2054278v1 *R. opacus* reference (Refseq assembly GCF_020542785.1), sorted and indexed. Expression counts were calculated for each genetic loci using featureCounts and imported into R for statistical analysis and visualization. Differential gene expression analysis was performed using DESeq2, pathway enrichment was performed using gage with the following parameters: test="unpaired", set.size = c(10,100), same.dir = TRUE, rank.test = FALSE, test4up=TRUE. Heatmaps were generated in R using pheatmap. Redundancy analysis, beta dispersion, and PERMANOVA were conducted using vegan in R.

4.2 Results

To identify transcription factors required for nitrogen-independent fatty acid production we utilized two previously published transcriptomic datasets to identify regulators differentially expressed (DE) in both nitrogen and phenolic stress. Of the 399 regulator genes reviewed, 141 responded with at least a two-fold change in transcription (either up- or down-regulation in comparison to cells cultivated in rich media) to phenol, and 216 responded to low nitrogen; this pool was narrowed to 33 candidates which responded to both stimuli, at least one of which was a strong (>2-fold) response (figure 4.1a).

These transcription factors were cloned into the high-copy pAL5000(S) backbone under a strong constitutive promoter; 30 plasmid constructs were confirmed by sequencing, and 27 were successfully overexpressed in PD630 (see **Error! Reference source not found.** for a full list). A final subset was selected for fatty acid fermentations in glucose and phenol (figure 4.1b).

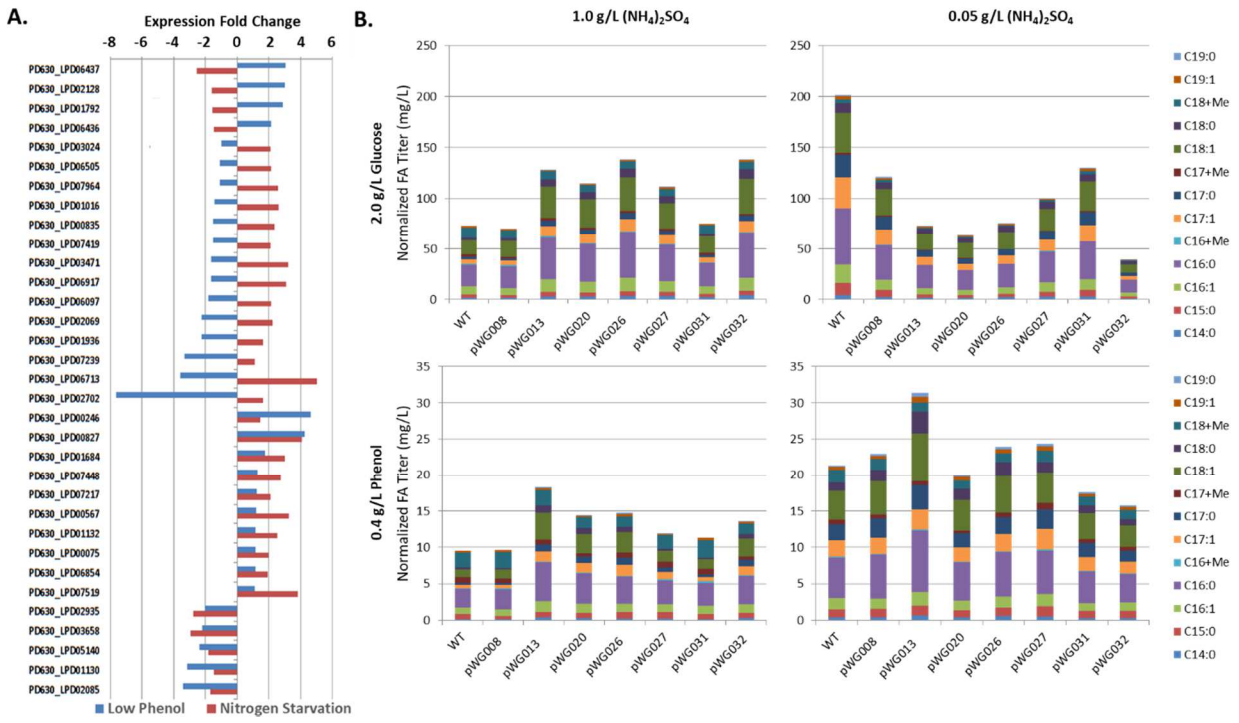


Figure 4. 1. ***Overexpressing autologous transcription factors with strong differential expression grown in phenol and nitrogen starvation results in increased triacylglyceride production in nitrogen rich environments***

a. Using two previous published datasets, transcriptional regulators which were differentially expressed (>1 log2 fold-change) in both phenol and during nitrogen starvation versus nutrient broth. 33 genes had greater than 2 log2 fold-change, and were selected for cloning. **b.** A subset of the 27 successfully overexpressed transcriptional regulators were selected for FA fermentation. Strains were grown in either glucose of phenol at two different nitrogen concentrations representing nitrogen rich- and nitrogen poor-environments for 60 hours, then lipids were extracted and measured using GC-FID. pWG013 and pWG026 produced higher lipid titers than WT in three conditions and was chosen along with pWG020 for transcriptomic analysis. All fatty acid titers have been normalized to an internal 40mg/L C12 standard and averaged across technical replicates.

4.2.1 Fatty acid accumulation in *R. opacus* is nitrogen concentration dependent when grown on glucose or phenol

In comparison to wild-type (WT) expressing an empty vector control, 5/7 characterized mutant strains produced higher titers of fatty acids in at least one of the four fermentation conditions – 2.0g/L glucose with either 1.0g/L or 0.05g/L $(NH_4)_2SO_4$, and 0.4g/L phenol with the previously mentioned concentrations of $(NH_4)_2SO_4$ (figure 1b). pWG013 and pWG026 increased fatty acid production versus WT in both nitrogen rich and poor conditions in phenol, as well as the nitrogen rich glucose condition pWG020, which produced higher titers of fatty acids than WT in two of four conditions (2.0g/L glucose with 1.0g/L $(NH_4)_2SO_4$ and 0.4g/L phenol with 1.0 g/L $(NH_4)_2SO_4$). Based on these measurements, we selected these three strains for transcriptomic analysis alongside WT.

4.2.2 Evidence of plasmid loss across multiple strains when grown in phenol and glucose

After demonstrating an increased accumulation of TAG in mutant strains grown on phenol and glucose, we conducted an analysis of the effect of autologous TR expression on the transcriptomic state of *R. opacus*. An initial assessment of expression coverage across the genome identified a complete lack of expression from all genes located on plasmid 3 in WT grown in both phenol and glucose, as well a majority of genes located on plasmid 2 in strains 13 and 20 grown in phenol and glucose (figure 4.6S). We thus removed those plasmids from analysis and concentrated our assessment of DE on loci present in all conditions.

4.2.3 ATR 13 and 20 overexpression induces transcriptional reprogramming in phenol and glucose

Using principal component analysis of the top 500 most highly variable genes, we leveraged redundancy analysis to regress strain identity and carbon source against the RNAseq

profiles of each sample (p = 0.0019 and 0.0005 for Strain and Condition, permuted (2000 times) fit to rda model, figure 4.2a).

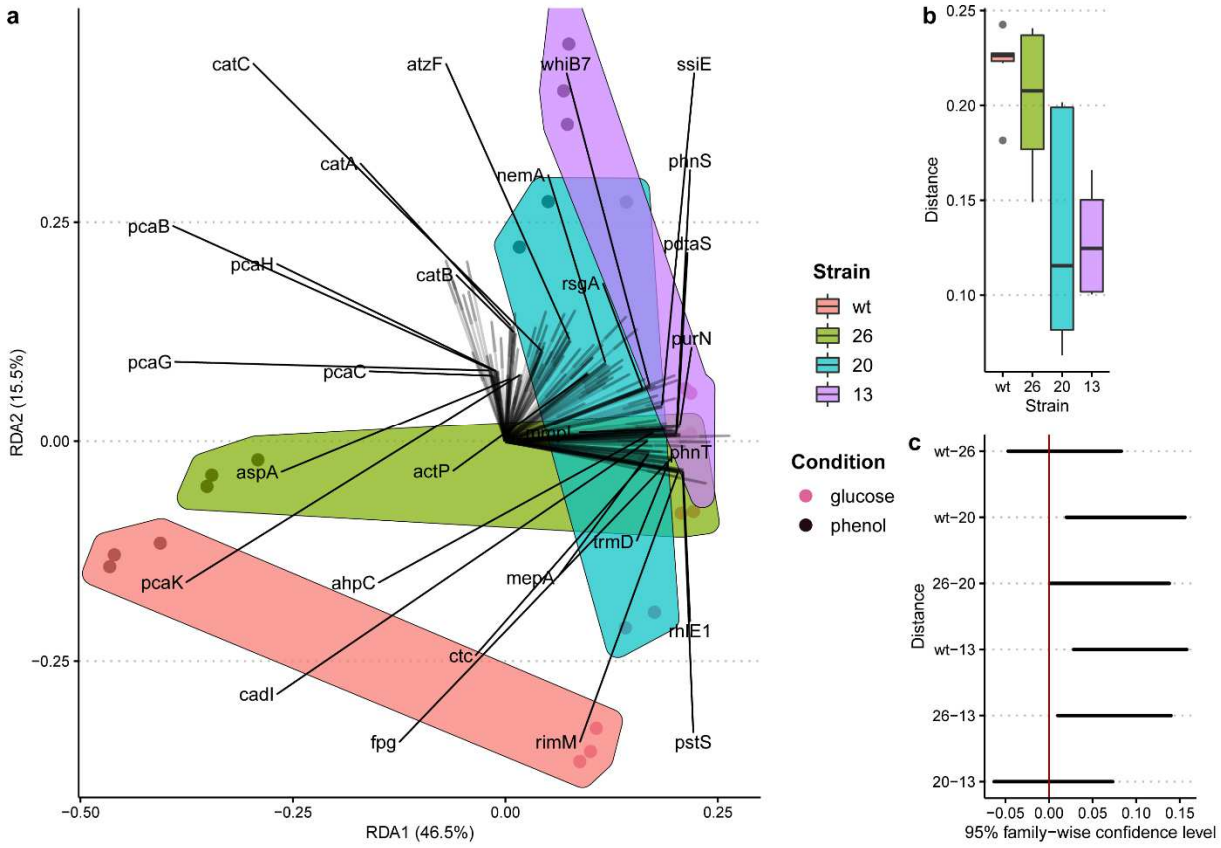


Figure 4. 2. ***ATR expression induces transcriptional reprogramming across strains and carbon sources***

a. RDA analysis was conducted on the first two principal component axes generated from the 500 most highly variable transcripts across all strains. Samples from each strain are illustrated by colored hulls. Samples from all strains grown in glucose have less variation than the samples grown in phenol, which vary farther across the two axes. Strains 13 and 20 also exhibit less dispersion between the samples grown in phenol vs the samples grown in glucose. Vector loadings which lie closest to the samples from strain 13 are drawn, and the transcripts with symbol annotations are displayed. **b.** Distance to centroid boxplots for each strain. The centroid of all the samples from a strain (across both conditions) was generated. WT had the highest distance between all samples, while Strain 13 had the lowest. **c.** 95% family-wise confidence level of the difference in dispersion between each strain. A vertical line demarcates zero; a

confidence interval which crosses this line indicates the true difference between the dispersion of the two conditions is likely zero. Only the WT-26 and 20-13 confidence levels cross 0.

Our initial examination of samples within PCA space detected large dispersion of the transcriptomes from all strains grown in phenol, and significantly lower distance to centroid for all samples grown in glucose versus phenol was confirmed (permutation test for homogeneity of multi-variate dispersions $p = 0.0002$). We next characterized the differences in dispersion between strains, as we noticed ellipses generated from strains 13 and 20 exhibited constriction when compared to WT and strain 26, which had far greater PCA distance between samples grown in each carbon source. Beta dispersion analysis identified differences between strains (permutation test for homogeneity of multi-variate dispersions $p = 0.0018$, figure 4.2b); a significantly lower distance to centroid for all samples within strains 13 and 20 versus WT and strain 26 was confirmed via a Tukey's HSD post-hoc test of the permuted dispersions (figure 4.2c). There were also no significant differences compositionally between strains 13 and 20 ($p = 0.1654$ – adonis2), indicating that ATR 13 and 20 overexpression produce transcriptomes which are both homogenous and compositionally similar. Interestingly, strain 26 exhibits similar homogeneity and composition to WT, ($p = 0.8608$ – Tukeys HSD and $p = 0.2803$ – adonis2), while still increasing TAG production.

4.2.4 ATR 13 and 26 increase expression of translation and phenol utilization genes

In order to identify components contributing to the altered transcriptomic state seen in 13 and 20 we identified the 100 loci loading magnitudes which were closest to the coordinates of the centroid generated from strain 13, the strain with the most significant difference in homogeneity and composition to WT ($p = 0.0035854$ – Tukeys HSD, figure 4.2b and 4.2c, and p

= 0.0024 – adonis2). The 32 annotated genes which explained the most variation in between those strains and the rest of the dataset were clustered into three distinct groups based on expression (figure 4.3):

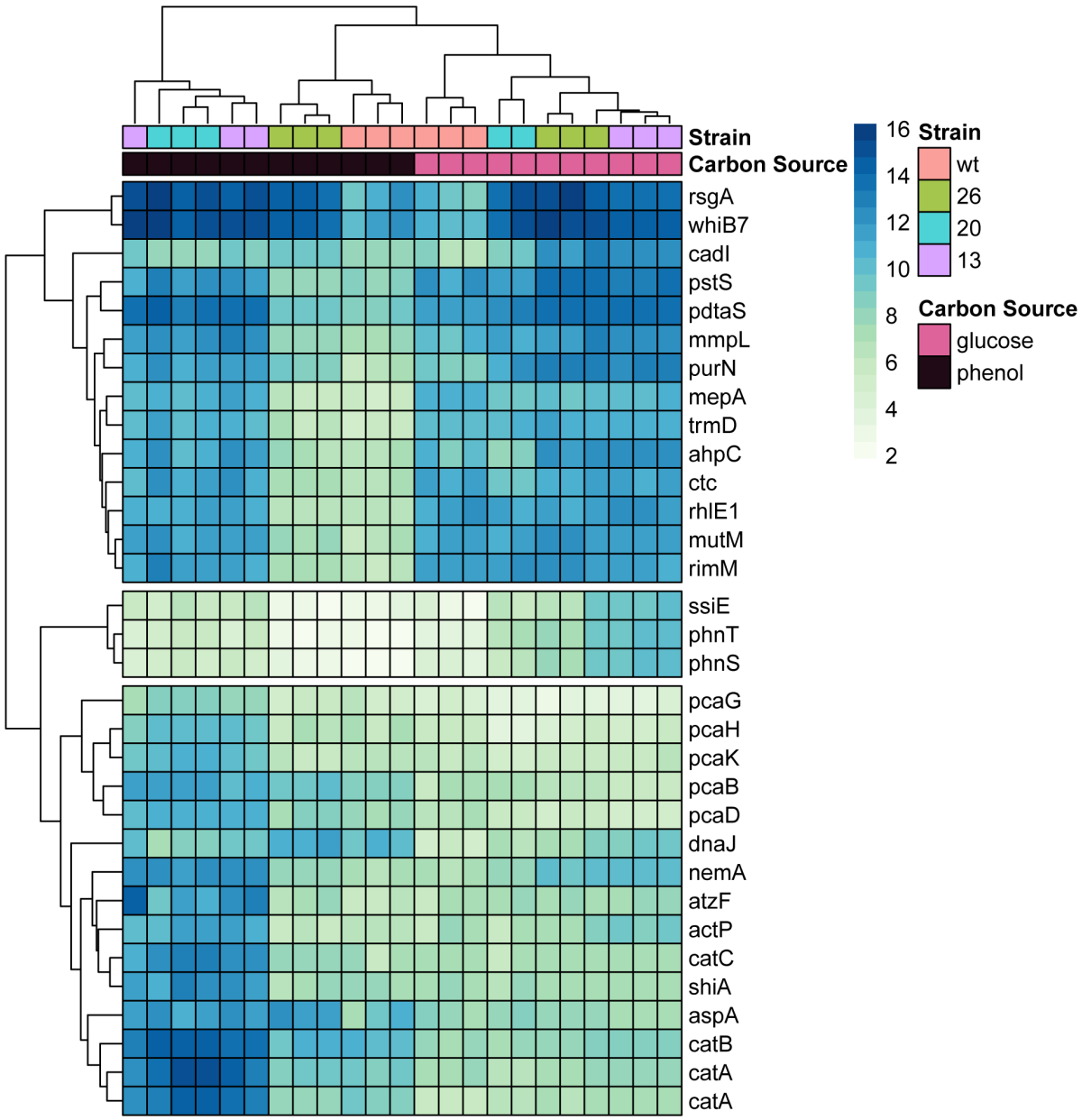


Figure 4. 3: *ATR 13 and 20 induce activation of carbon-specific alternate transcriptional programs*

a. a heatmap of the loci with RDA loadings lying closest to Strain 13 and 20. Column annotations indicate the strain identity and carbon source the samples were grown in. Rows are clustered into 3 groups: 1. The "glucose optimization gene set for TAG production" or GOG, which are genes upregulated in glucose in WT which are become universally upregulated by Strain 13 and 20 in either glucose or phenol, 2. A set of three genes which are only upregulated in Strain 13 grown in phenol. and 3. The "phenol optimization gene set for TAG production" or POG, which are genes upregulated in strain 13 and 20 versus wt.

1. A set of genes absent in WT and 26 grown in phenol but expressed in strains 13 and 20 regardless of carbon source (hereafter referred to as "carbon source universal optimization gene set for TAG production", or COG), 2. A set of 3 genes (the transport genes *phnT* and *phnS*, and the biofilm promoting virulence factor *sslE(18)*) that were only upregulated in strain 13 grown in glucose, and 3: a set of genes differentially expressed in strains 13 and 20 versus WT grown in phenol (hereafter referred to as "phenol-specific optimization gene set for tag production", or POG). COG contains a set of genes involved in translation (*rhlE1, cpc, rimM,* and *trmD (19, 20)*), and *pstS* encoding phosphate transport(21). Intrestingly, *pdtaS*, the sensory histidine kinase of the *pdtaR/S* two-component regulatory system was also a member of this set of genes(22). A subset of this third set, (*rsgA, purN, phnT/S, pdtaS, pstS,* and *ahpC)* was significantly upregulated in Strain 13 when grown on glucose versus WT. Among the genes with only basic annotations, gtrA-like glucose translocase K2Z90_RS09460 was upregulated in strains 13, 20, and 26 grown in glucose. The POG contained a set of genes highly upregulated in Strain 13 and 20 compared to WT. This included *catABC* gene cluster of the ortho cleavage arm of catechol degradation(23), the *pcaBDHGK* genes of the protocatechuate degradation II pathway important for degradation of aromatic carbon sources(24), and the aromatic amino acid biosynthesis gene shikimate transporter shiA(25).

4.2.5 The PaaX-like gene ATR 13 induces upregulation of metabolic pathways related to protein and cofactor synthesis in phenol

ATR 13 is annotated as a PaaX-family transcriptional regulator. As stated previously, when overexpressed ATR 13 increases the compositional similarity between the transcriptional state of Strain 13 when grown in glucose or phenol as a sole carbon source. We conducted KEGG Pathway analysis to determine which cellular components underwent differential expression, comparing all strains in both carbon sources to WT (figure 4.4a).

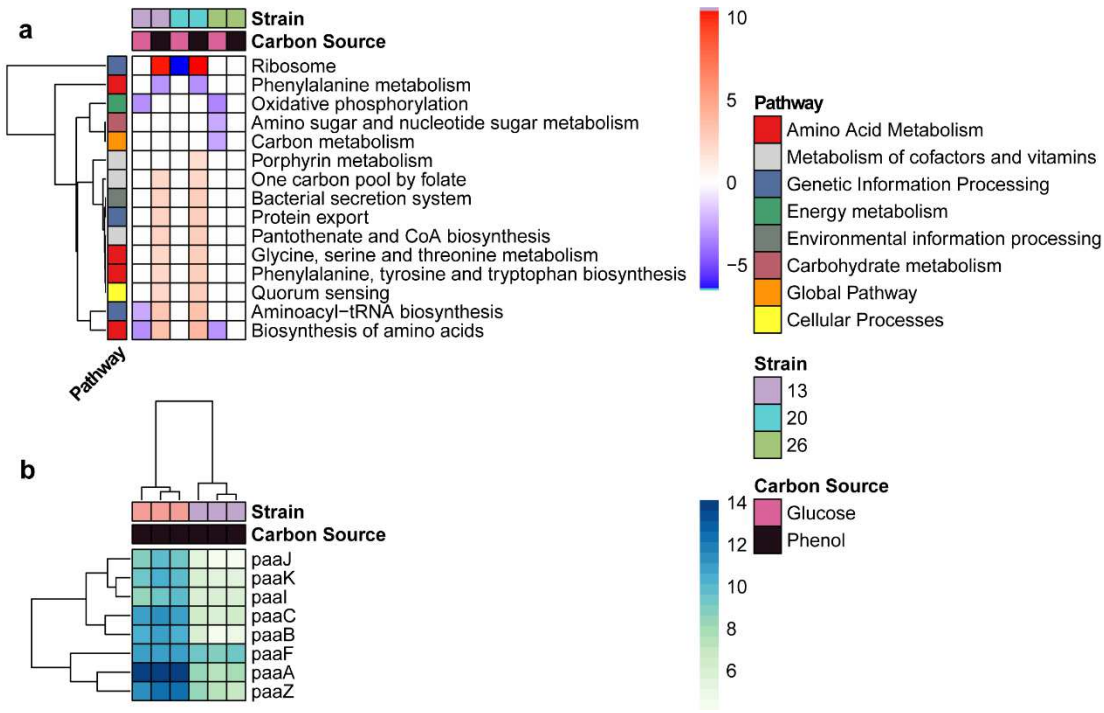


Figure 4. 4: ***The PaaX-like gene ATR 13 induces carbon-specific alternate transcriptional programs and represses phenylacetic acid degradation***

a. Heatmap visualizing the DE of KEGG pathways. Each mutant condition was compared to the WT condition grown at the same carbon source for differential expression to all the *R. opacus* annotated KEGG pathways (rows). Color of each cell of the heatmap denotes the fold change versus WT, and white cells represent non-significant changes. Each column represents the DE of

one condition, averaged from replicates and tested using GAGE in R. Annotation bars on top of the heatmap denote the strain and carbon source each column, and the row annotation bar on the side denotes the BRITE functional hierarchy classification for each pathway. Rows are clustered using the "complete" distance method. **b.** Heatmap visualizing the DE of the phenylacetic acid degradation pathway in strain 13 versus WT grown on phenol. Here, all loci annotated to the phenylacetic acid pathway were shown as rows, and each column represents one of the replicates for each condition. Column annotation bars on top of the metric indicate the strain and carbon source designation of each replicate. Rows and columns are clustered using the standard pheatmap distance and clustering metric.

ATR 13 downregulated 3 KEGG pathways when grown in glucose: oxidative phosphorylation, aminoacyl-tRNA biosynthesis, and the biosynthesis of amino acids, while ATR 20 only downregulated the ribosome. ATR 13 grown in phenol induced increased expression 10 KEGG pathways, including pantothenate and coa metabolism, phenylalanine, tyrosine and tryptophan biosynthesis, ribosome, and aminoacyl-tRNA synthesis. ATR 20 upregulated the same set of 10 pathways, but also upregulated porphyrin metabolism. Both ATR 13 and ATR 20 upregulated multiple metabolism and cofactor pathways as well as pathways related to amino acid metabolism. Strain 26 did not upregulate any pathways in either carbon condition, and instead downregulated 4 in glucose: oxidative phosphorylation, amino and nucleotide sugar biosynthesis, carbon metabolism, and biosynthesis of amino acids. Strain 26 did not differentially express any pathways versus WT.

To confirm ATR 13's role in regulating the paa pathway, we looked for differential expression at the loci annotated to be a part of the paa cluster (figure 4.4b). All genes within the cluster were significantly downregulated in Strain 13 grown in phenol compared to WT. Differential expression of KEGG modules within the phenylalanine, tyrosine, and tryptophan metabolism and phenylalanine degradation pathways further identified the "Phenylacetate degradation, phenylacetate to acetyl-CoA/succinyl-CoA" module as being significantly downregulated in all

mutant strains grown in phenol (Table S2). All three strains also exhibited upregulation of the "Tryptophan biosynthesis, chorismate to tryptophan". Strains 13 and 20, but not 26, also upregulated the "Shikimate pathway, phosphoenolpyruvate + erythrose-4P to chorismate" module.

4.2.6 The PaaX-like gene ATR 13 requires feaR, the phenylethylamine degradation pathway transcriptional regulator, to increase FA titers

PaaX is a negative regulatory protein acting on the phenylacetic acid degradation operon (paa), or Paa gene cluster. There are 8 annotated *paa* cluster genes in *R. opacus*. Additionally, a feaR transcription factor and a 6-aminohexanoate-cyclic-dimer hydrolase clustered together with the PAA catabolism operon (Fig. 4.5a). To determine whether a non-functioning *paa* pathway was required for increased FA titers in strain 13, the first gene in the PAA catabolism pathway, phenylacetate-coenzyme A ligase, was selected and knocked out via homologous recombination. This resulted in loss of increased FA titer, returning to WT levels. (Fig. 4.5b – glucose and 4.5c – phenol), indicating that complete loss of the paa pathway results in loss of increased FA production. *feaR,* a transcriptional regulator responsible for activation of the phenylethylamine degradation pathway is located directly adjacent to the paa cluster. To test whether loss of phenylethylamine degradation also results in decreased FA titers, we deleted *feaR* in Strain 13 (07015, figure 4.5a). Indeed, this resulted in loss of the increased lipid production phenotype. In contrast, deleting 07014 a 6-aminohexanoate-cyclic-dimer hydrolase located near the paa cluster was not required for increased FA titers. To further confirm the potential role of the *feaR* in regulating lipid biosynthesis in R. opacus, we overexpressed this transcriptional regulator in the WT strain. Lipid profile analysis revealed that in nitrogen replete conditions with either glucose

or phenol as carbon source, the lipid titer was increased by 36.6% and 56.2%, respectively (Fig. 4.5d,e, p value < 0.05, Student's t test).

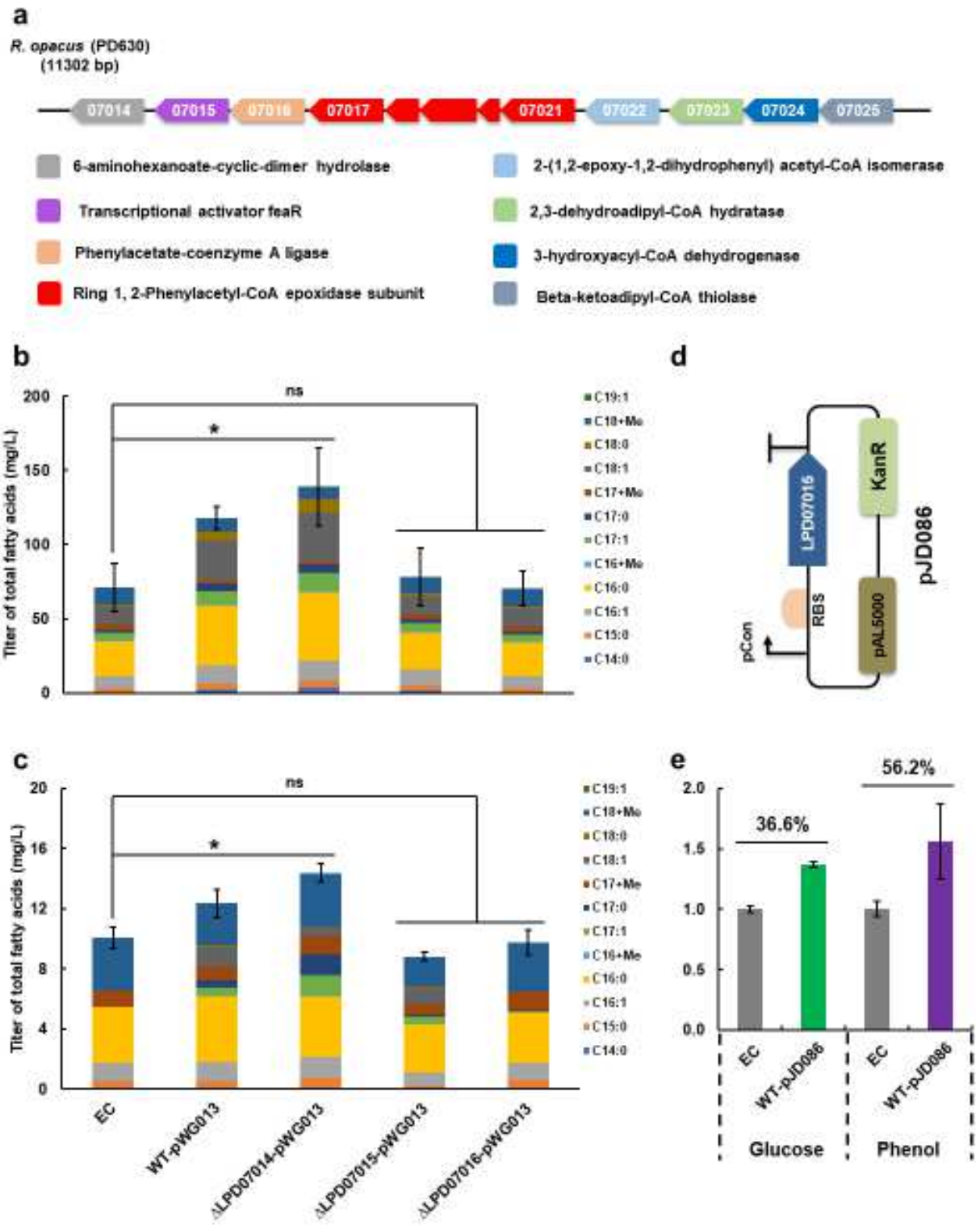


Figure 4. 5. ***Mining the downstream target of* paaX.** ***a****, The proposed phenylacetic acid (PAA) degradation operon. Genes are shown with LPD gene numbers from the NCBI database (Refseq, CP003949.1).* ***b****,* ***c****,*

*Assays of lipid profiles of different strains with cultivation on 2 g/L glucose (**b**) and 0.4 g/L phenol (**c**) as carbon source. The WT and the mutant strains harbor the transcription factor* paaX *overexpression plasmid pWG013. The WT strain which harbors the empty vector was used as a control (EC). **d**, Schematic diagram of the replicating plasmid used for overexpressing the feaR transcription factor. **e**, The changes of the lipid production in the strain harbors pJD086. The control (EC) and the engineered strains were cultivated on 2 g/L glucose and 0.4 g/L phenol, respectively. All the assays were conducted on the nitrogen replete condition (1g/L ammonium sulfate). All values represent the mean of triplicate cultures, with error bars depicting the standard deviation from that mean. Unpaired two-tailed t-test was used to compare the variation in the change of lipid contents of the mutants against that of the EC strain (*, P <0.05, ns, not significant).*

4.4 Discussion

Carbon storage in the bacterial cell is usually induced during nutrient limitation (phosphorus, nitrogen, etc.) and results in slowed growth. Genetic components involved in modulating TAG synthesis in *R. opacus PD630* have been discovered(11, 12), but an investigation into putative regulatory elements necessary for TAG accumulation in phenol has hitherto not been attempted. In this study, we observe increases in TAG production through overexpression of autologous transcriptional regulators, describe the effects of these modulations on the transcriptome of the cell chassis, and use experimental techniques to interrogate this mechanism further.

Of the 7 ATR mutants selected for FA titer measurement, ATR 13, 20, and 26 exhibited consistently higher FA titers across all four conditions (figure 1). Conversely, ATR 08 and ATR 31 overexpression consistently resulted in minimal changes to FA titers. While this is not necessarily surprising for ATR 08, which is annotated as yobV, an uncharacterized HTH-type transcriptional regulator, ATR 31 encodes the nitrate regulatory gene narL, known to play a role in regulating activity of terminal electron receptors depending on environmental cues(26). This suggests that mechanisms for severing the relationship between FA accumulation and nitrogen availability are only tangentially related to the classical nitrogen regulatory systems. This

observation is further supported by observing the functions of ATR 13, 20, and 26, which consistently increased FA titers, being orthogonal to nitrogen resource regulation in *R. opacus*.

ATR expression in glucose resulted in significantly less compositional change across all strains, suggesting the mechanisms used for increasing TAGs are already active during glucose degradation. The COGs correlating most strongly with strain 13 grown in glucose were primarily related to translation (*rhlE1, cpc, rimM,* and *trmD*), likely due to the increased synthesis demand during protein overexpression. ATR13 and ATR20 successfully induce this gene set during growth on phenol as well, resulting in Strain 13 exhibiting no significant differential expression in KEGG pathways when grown in phenol versus glucose. The indicates that the mechanism by which Strain 13 increases TAG production results in a similar transcriptional state regardless of the carbon source.

The fatty acid biosynthesis and triacyglyceride biosynthesis pathways were not differentially regulated. Several genes associated with increased TAG production have been reported previously in *Rhodococcus* species. The major lipid droplet protein tadA(27), required for lipid storage and droplet formation, was significantly upregulated in Strains 13, 20, and to a lesser extent, in Strain 26, as well as MLDSR, the regulator which controls tadA expression(28). Overproduction of fadR, the fatty acid biosynthesis promoter, leads to abnormally large cell size and decreased cell growth(29), while overexpression of *accB* or *accC* led to inhibition of the *bio* operon required for biotinylation of acetyl-Coa carboxylase, the first step of lipid synthesis(30). Overexpression of coenzyme A may thus represent a method of increasing TAGs without some of the negative effects of direct modulation of the aforementioned pathways. Coenzyme A overexpression to increase TAG biosynthesis in oleaginous bacteria has already been successfully implemented in algae(31).

Though we profiled three ATR mutants with increased levels of TAG production, we observed the creation of only two unique transcriptional states. ATR 26, which produces a statistically similar level of dispersion and composition to WT *R. opacus PD630*, induces a markedly different transcriptional regime than ATR 13 and 20 and little to no pathway DE versus WT. It is unclear whether unique transcriptomes for increased TAG production in *R. opacus* is indicative of multiple mechanisms for altering storage accumulation, or if ATR 13 and 20 represent effectors farther upstream of the same mechanism, control larger regulons, or produce increased pleiotropic effects by some other method. Further work interrogating the mechanism by which ATR 26 overexpression increases TAG bioproduction is necessary to determine which of these two possibilities is true.

ATR 13 increased FA titers in three of the four experimental conditions, improving utilization in phenol regardless of nitrogen concentration. The compositional changes to expression profiles during growth on phenol for Strains 13 are directly tied to increased expression of phenol catabolism genes. Both the *pca* and *cat* branches of the beta-ketoadipate pathway were included in the POGs increased in expression in strains 13 and 20 versus WT. Though previous work has shown phenol to be utilized solely through the *cat* branch(9), upregulation of both branches of the B-ketoadipate pathway was also seen in WT and in *R. opacus PD630* strains adapted to growth on multiple aromatic compounds(9). *E. coli paaJ*, a thiolase gene within the phenylacetic acid degradation pathway repressed in Strain 13, exhibits high sequence similarity and) to *PcaF* (66.1%, *P. putida*) and *CatF* (61.3%, *A. calcoaceticus*) of the protocatechuate and catechol arms of the beta-ketoadipate pathway(32). Repression of the phenylacetate degradation pathway, which produces similar product intermediates, compete for the same cofactors and cellular resources, and shares high protein identity, results in upregulation

of the primary phenol catabolism pathway in *R. opacus.* This could lead to higher phenol utilization efficiency as a sole carbon source to produce the important cellular components acetyl-CoA and succinyl-CoA – moreover, microbial engineering to reduce cofactor demand has been shown to be an effective method of increasing flux through desired metabolic pathways(33, 34). *feaR* encodes for the activator of the phenylethylamine degradation pathway, which produces phenylacetate(35). The importance of upstream branches of the phenylalanine degradation pathway was demonstrated by loss of TAG production in a Δ*feaR/ΔPaaX* background.

These experiments uncovered evidence linking increased production of TAG with alterations to the expression of phenylalanine metabolism and CoA biosynthesis. ATR 13 is annotated as a PaaX-like protein involved in the regulation of paa degradation. ATR 13 and 20 overexpression both lead to repression of the *paa* gene cluster. As strains 20 and 13 had similar levels of beta-dispersion across carbon compounds and no significant differences in transcriptome composition, we can conclude that ATR 20 regulates TAG production through a similar mechanism. There is growing evidence that phenylacetate is a cross-kingdom signaling molecule with important contributions to the oxidative stress response in *Acinetobacter baumannii,* and is known to regulate growth and development in plants(36). Given that complete abrogation of phenylacetate synthesis by deleting genes within the *paa* gene cluster resulted in loss of increased FA titers; it seems likely that some paa expression is needed within the cell, perhaps to allow *R. opacus PD630* to efficiently respond to the lipid peroxidation associated with increased oxidative stress. ATR 13 grown in phenol increased phenylalanine, tyrosine and tryptophan biosynthesis expression, resulting in upregulation of the shikimate degradation and tryptophan biosynthesis pathway modules, indicating chorismate is shunted away from

phenylacetic acid degradation and towards other products. Bacterial degradation of tyrosine produces primarily phenol(37), and ATR 13 and 20 overexpression may produce more substrate for the beta-ketoadipate pathway. Taken together, these data suggest that selectively tuning the expression of the paa pathway shunts cofactor resources towards phenol degradation and other aromatic metabolism pathways, resulting in increased TAG production.

In conclusion, we have demonstrated the ability to increase TAG production through autologous transcriptional regulator overexpression. The effects on the transcriptome were significantly greater when grown in phenol. Experimental validation revealed the necessity of the upper branches of the phenylalanine metabolism resulting in the production of phenylacetate. Selective tuning of the paa pathway resulted in the increases in several cofactor pathways resulting in increased phenol catabolism, likely due to re-allocation of cellular resources. This work describes a novel mechanism for increasing TAG production in phenol and glucose.

4.5 Acknowledgements

This work was created in collaboration with Rhiannon Carr, Weitao Geng, Jinjin Daio, Tae-Seok Moon, Gautam Dantas, and Fuzhong Zhang. *In preparation*. F.Z., and W.G., J.D, W.E.A., T.S.M and G.D. conceived the project. W.G. engineered the microbial strains. W.G., R.C, and J.D. performed fermentation, and FA analysis; W.G., R.C, and J.D. performed the lipid analysis; W.G., R.C, and J.D. performed phenotypic profiling and growth experiments; W.E.A. performed the transcriptomic analysis. All authors analyzed the data and wrote the paper. This work was supported by the Department of Energy (DESC0018324).

4.5 Appendix (Supplementary Material)

4.5.1 Tables

Table 4. 1. Overexpression strains screened for nitrogen-independent fatty acid production in Rhodococcus opacus PD630.

Overexpression Strain	Locus ID	Gene Name
PD630 + pWG001	LPD06713	yagI
PD630 + pWG002	LPD02702	liaR
PD630 + pWG005	LPD01792	degU
PD630 + pWG006	LPD06436	dosT
PD630 + pWG008	LPD02128	yobV
PD630 + pWG009	LPD07239	ybhD
PD630 + pWG012	LPD06097	BetI
PD630 + pWG013	LPD06917	paaX
PD630 + pWG014	LPD03471	yidP
PD630 + pWG015	LPD07419	ycbG
PD630 + pWG016	LPD00835	ttgR
PD630 + pWG018	LPD07964	yiaJ
PD630 + pWG019	LPD06505	kipR
PD630 + pWG020	LPD3024	Rv0472c-MT0489
PD630 + pWG024	LPD00827	ttuE_C
PD630 + pWG025	LPD01684	yxaF
PD630 + pWG026	LPD00567	tcmR
PD630 + pWG027	LPD07519	slyA
PD630 + pWG028	LPD02935	Leu

PD630 + pWG029	LPD03658	HI_1364
PD630 + pWG030	LPD01130	pro_trans
PD630 + pWG031	LPD02085	NarL
PD630 + pWG032	LPD07217	ttgR
PD630 + pWG033	LPD05140	Mb0601
PD630 + pWG034	LPD01132	yagI
PD630 + pWG035	LPD00075	ribonuclease
PD630 + pWG036	LPD06854	famr

4.5.1 Figures

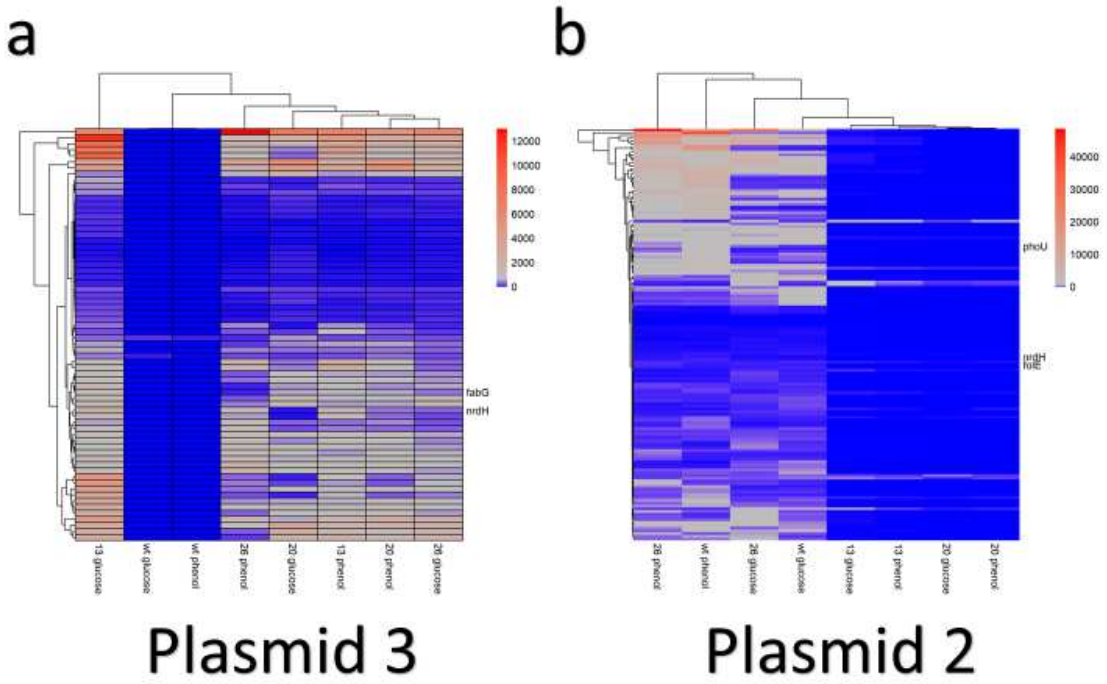


Figure 4. 6S ***Potential plasmid loss in multiple strains on growth in phenol and glucose.***

a. Raw transcript counts of all genes located on plasmid 3. Overall, though there is one area of low coverage, there is expression across most of the plasmid for all strains except WT grown in phenol. Symbols of annotated genes appear along the right side of the heatmap. b. Raw transcript

counts of all genes located on plasmid 2. Like plasmid 2, most strains express parts of the plasmid, except for strains 13 and 20 grown in either glucose or phenol. Symbols of annotated genes appear along the right side of the heatmap.

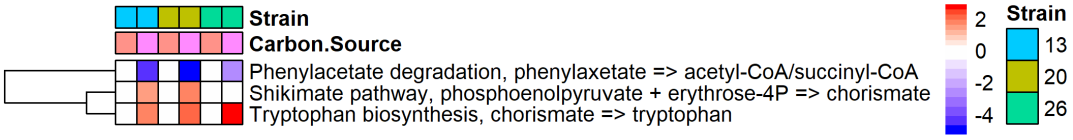


*Figure 4. 7**s KEGG modules within the Phenylalanine, tyrosine, and tryptophan biosynthesis and phenylalanine metabolism pathways with significant DE***

Heatmap of KEGG module enrichment in each strain and carbon source. Each mutant condition was compared to the WT condition grown at the same carbon source for differential expression to all the *R. opacus* annotated KEGG modules (rows). Color of each cell of the heatmap denotes the fold change versus WT, and white cells represent non-significant changes. Each column represents the DE of one condition, averaged from replicates and tested using GAGE in R. Annotation bars on top of the heatmap denote the strain and carbon source each column. Rows are clustered with the "complete" distance method using minkoswki distance.

4.6 References

1. Harkin JM. LIGNIN AND ITS USES. FOREST PRODUCTS LAB MADISON WIS; 1969.

2. Beckham GT, Johnson CW, Karp EM, Salvachúa D, Vardon DR. Opportunities and challenges in biological lignin valorization. Current Opinion in Biotechnology. 2016;42:40-53.

3. Patton AR, Gieseker L. Seasonal Changes in the Lignin and Cellulose Content of Some Montana Grasses. Journal of Animal Science. 1942;1(1):22-6.

4. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. Lignin Biosynthesis and Structure. Plant Physiology. 2010;153(3):895-905.

5. Mottiar Y, Vanholme R, Boerjan W, Ralph J, Mansfield SD. Designer lignins: harnessing the plasticity of lignification. Current Opinion in Biotechnology. 2016;37:190-200.

6. Holder JW, Ulrich JC, DeBono AC, Godfrey PA, Desjardins CA, Zucker J, et al. Comparative and Functional Genomics of Rhodococcus opacus PD630 for Biofuels Development. PLOS Genetics. 2011;7(9):e1002219.

7. Alvarez H, Steinbüchel A. Triacylglycerols in prokaryotic microorganisms. Applied Microbiology and Biotechnology. 2002;60(4):367-76.

8. Alvarez HM, Mayer F, Fabritius D, Steinbüchel A. Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630. Archives of Microbiology. 1996;165(6):377-86.

9. Henson WR, Campbell T, DeLorenzo DM, Gao Y, Berla B, Kim SJ, et al. Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. Metabolic engineering. 2018;49:69-83.

10. Subramaniam R, Dufreche S, Zappi M, Bajpai R. Microbial lipids from renewable resources: production and characterization. Journal of Industrial Microbiology and Biotechnology. 2010;37(12):1271-87.

11. Hernandez MA, Arabolaza A, Rodriguez E, Gramajo H, Alvarez HM. The atf2 gene is involved in triacylglycerol biosynthesis and accumulation in the oleaginous Rhodococcus opacus PD630. Appl Microbiol Biotechnol. 2013;97(5):2119-30.

12. Alvarez AF, Alvarez HM, Kalscheuer R, Wältermann M, Steinbüchel A. Cloning and characterization of a gene involved in triacylglycerol biosynthesis and identification of additional homologous genes in the oleaginous bacterium Rhodococcus opacus PD630. 2008;154(8):2327-35.

13. Alvarez HM, Steinbüchel A. Biology of Triacylglycerol Accumulation by Rhodococcus. In: Alvarez HM, editor. Biology of Rhodococcus. Cham: Springer International Publishing; 2019. p. 299-332.

14. Xie S, Sun S, Lin F, Li M, Pu Y, Cheng Y, et al. Mechanism-Guided Design of Highly Efficient Protein Secretion and Lipid Conversion for Biomanufacturing and Biorefining. Advanced Science. 2019;6(13):1801980.

15. Hernández MA, Alvarez HM. Increasing lipid production using an NADP(+)-dependent malic enzyme from Rhodococcus jostii. Microbiology (Reading). 2019;165(1):4-14.

16. Huang L, Zhao L, Zan X, Song Y, Ratledge C. Boosting fatty acid synthesis in Rhodococcus opacus PD630 by overexpression of autologous thioesterases. Biotechnology Letters. 2016;38(6):999-1008.

17. Villalba MS, Alvarez HM. Identification of a novel ATP-binding cassette transporter involved in long-chain fatty acid import and its role in triacylglycerol accumulation in Rhodococcus jostii RHA1. Microbiology. 2014;160(7):1523-32.

18. Decanio MS, Landick R, Haft RJ. The non-pathogenic Escherichia coli strain W secretes SslE via the virulence-associated type II secretion system beta. BMC Microbiol. 2013;13:130.

19. Kaczanowska M, Rydén-Aulin M. Ribosome Biogenesis and the Translation Process in Escherichia coli. Microbiology and Molecular Biology Reviews. 2007;71(3):477-94.

20. Jain C. The E. coli RhlE RNA helicase regulates the function of related RNA helicases during ribosome assembly. Rna. 2008;14(2):381-9.

21. Díaz M, Esteban A, Fernández-Abalos JM, Santamaría RI. The high-affinity phosphate-binding protein PstS is accumulated under high fructose concentrations and mutation of the corresponding gene affects differentiation in Streptomyces lividans. Microbiology (Reading). 2005;151(Pt 8):2583-92.

22. Morth JP, Gosmann S, Nowak E, Tucker PA. A novel two-component system found in Mycobacterium tuberculosis. FEBS Letters. 2005;579(19):4145-8.

23. Nešvera J, Rucká L, Pátek M. Chapter Four - Catabolism of Phenol and Its Derivatives in Bacteria: Genes, Their Regulation, and Use in the Biodegradation of Toxic Pollutants. In: Sariaslani S, Gadd GM, editors. Advances in Applied Microbiology. 93: Academic Press; 2015. p. 107-60.

24. Li C, Zhang C, Song G, Liu H, Sheng G, Ding Z, et al. Characterization of a protocatechuate catabolic gene cluster in Rhodococcus ruber OA1 involved in naphthalene degradation. Annals of Microbiology. 2016;66(1):469-78.

25. Whipp MJ, Camakaris H, Pittard AJ. Cloning and analysis of the shiA gene, which encodes the shikimate transport system of escherichia coli K-12. Gene. 1998;209(1-2):185-92.

26. Iuchi S, Lin EC. The narL gene product activates the nitrate reductase operon and represses the fumarate reductase and trimethylamine N-oxide reductase operons in Escherichia coli. Proc Natl Acad Sci U S A. 1987;84(11):3901-5.

27. Ding Y, Yang L, Zhang S, Wang Y, Du Y, Pu J, et al. Identification of the major functional proteins of prokaryotic lipid droplets. J Lipid Res. 2012;53(3):399-411.

28. Zhang C, Yang L, Ding Y, Wang Y, Lan L, Ma Q, et al. Bacterial lipid droplets bind to DNA via an intermediary protein that enhances survival under stress. Nat Commun. 2017;8:15979.

29. Vadia S, Tse JL, Lucena R, Yang Z, Kellogg DR, Wang JD, et al. Fatty Acid Availability Sets Cell Envelope Capacity and Dictates Microbial Cell Size. Curr Biol. 2017;27(12):1757-67.e5.

30. Abdel-Hamid AM, Cronan JE. Coordinate expression of the acetyl coenzyme A carboxylase genes, accB and accC, is necessary for normal regulation of biotin synthesis in Escherichia coli. J Bacteriol. 2007;189(2):369-76.

31. Avidan O, Brandis A, Rogachev I, Pick U. Enhanced acetyl-CoA production is associated with increased triglyceride accumulation in the green alga Chlorella desiccata. Journal of Experimental Botany. 2015;66(13):3725-35.

32. Ferrández A, Miñambres B, García B, Olivera EaR, Luengo JM, García JL, et al. Catabolism of Phenylacetic Acid in Escherichia coli: CHARACTERIZATION OF A NEW AEROBIC HYBRID PATHWAY*. Journal of Biological Chemistry. 1998;273(40):25974-86.

33. Akhtar MK, Jones PR. Cofactor Engineering for Enhancing the Flux of Metabolic Pathways. Frontiers in Bioengineering and Biotechnology. 2014;2.

34. Li N, Zeng W, Xu S, Zhou J. Toward fine-tuned metabolic networks in industrial microorganisms. Synthetic and Systems Biotechnology. 2020;5(2):81-91.

35. Zeng J, Spiro S. Finely tuned regulation of the aromatic amine degradation pathway in Escherichia coli. J Bacteriol. 2013;195(22):5141-50.

36. Hooppaw AJ, McGuffey JC, Venanzio GD, Ortiz-Marquez JC, Weber BS, Lightly TJ, et al. The Phenylacetic Acid Catabolic Pathway Regulates Antibiotic and Oxidative Stress Responses in Acinetobacter. mBio. 2022;13(3):e01863-21.

37. Smith EA, Macfarlane GT. Dissimilatory Amino Acid Metabolism in Human Colonic Bacteria. Anaerobe. 1997;3(5):327-37.

Chapter 5:

Adaption to model lignin breakdown compounds results in divergent evolutionary trajectories

5.1 Introduction

5.1.1 Abstract

Biofuels have the potential to reduce carbon emissions, but currently only represent a small portion of transportation energy consumption. Utilization of the lignin fraction of lignocellulosic biomass could increase the efficiency of biofuel production and speed further adoption. Hybrid depolymerization techniques reducing lignin to simple breakdown products for microbial degradation have been proposed, and *Rhodococcus opacus PD630* is a promising chassis for this application. In previous work, *R. opacus* strains were evolved for increased tolerance to aromatic compounds. Here we phenotypically characterize each strain for changes to growth, carbon source utilization, and transcriptional expression as a result of adaption and increased aromatic concentration. We identify increased utilization of vanillic acid in adapted strains, and concentration dependent preferential utilization of 4-hydroxybenzoic acid at concentrations non-permissive to WT growth. While there is minimal differential expression of the aromatic catabolism pathways in adapted mutants at low concentrations, we observe divergent expression of the β-ketoadipate pathway in adapted mutants, and a similar expression pattern in a novel putative operon likely involved in aromatic catabolism.

5.1.2 Introduction

Modern, globalized society depends on a vast network of planes, trains, and automobiles in order to transport people and products around the earth. This accounts for 28% of all U.S. energy consumption, 90% of which is fueled by petrochemicals(1). The United States of America, the number 2 source of carbon emission by country in the world, has committed to net zero emissions by 2050(2). In support of this goal, the Agricultural Innovation Agenda, a USDA department wide effort to reduce carbon emissions and boost the use of renewable fuel sources, has further committed to increasing biofuel feedstock production and achieving a biofuel blend rate of E30 by 2050(3). This will require a substantial increase in production and usage; as of 2022 biofuels represent only 6% of total transportation(1). Production of bioethanol is primarily hampered by inefficiency: lignin, which encompasses 10-25% of woody tissue mass in lignocellulosic biomass, is currently not utilized as a biofuel precursor, and instead often burned after collection(4, 5). This is primarily because lignin is a complex, heterogeneous molecule derived from multiple cross-linked monolignol alcohol groups, primarily *p*-coumaryl, coniferyl, and sinapyl(6). The actual composition of any particular plant stock-derived lignin is hard to predict and varies from species-to-species and by season(6-8). As such, degradation of lignin requires an energy-intensive pre-treatment process generally yielding aromatic and phenolic derivatives(9-12). This extensive treatment necessary, usually involving harsh treatment chemicals and energy, has kept lignin utilization from commercial viability due to excessive cost(11, 13). Hybrid depolymerization techniques involving thermochemical pre-treatment followed by further depolymerization using biological agents have been suggested as a method of lowering utilization costs(13, 14).

Rhodococcus opacus PD630 (hereafter referred to as *R. opacus*), is particularly suited for use in the bio-chemical conversion of lignin due to its ability to tolerate inhibitory compounds(15-18). Degradation gene clusters for many monoaromatics derived from lignin (hereafter referred to as model lignin breakdown products, or MLBPs), have been identified, as well as the conversion paths of each monomer through the β-ketoadipate pathway(17, 19). There have been attempts to adapt *R. opacus spp.* for increased tolerance and bioproduction on lignin(20). In one study, *R. opacus MITXM-61*, a close relative of *PD630,* was adaptively evolved on multiple lignin derived compounds, such as phenol, vanillin, as well as lignin itself(16). However no genetic characterization to identify causal mechanisms for tolerance occurred. Another group utilized a novel gene deletion/insertion system to engineer *R. opacus* for increased production of *cis-cis* muconate through interruption of the β-ketoadipate pathway, and successfully accumulated bioproduct from phenol, vanillic acid, p-hydroxybenzoate (4-hydroxybenzoate) and lignin(21). Interruption of the β-ketoadipate pathway unfortunately resulted in increased sensitivity to these aromatic compounds and lignin, resulting in longer lag time and lower biomass. The low yield from adapted *R. opacus* grown on lignin, vanillin, and p-hydroxybenzoate suggests that along with the need for a better understanding of the mechanisms for increased tolerance to substrate toxicity, unknown pathways for aromatic utilization exist which have yet been described.

Due to its oleaginous nature, *R. opacus* produces triacylglyceride (TAG) carbon storage molecules instead of the more common polyhydroxyalkanoic acid, as much as ~44% of dry cell weight (DCW) when grown on aromatic carbon sources(19) and lignin deplymerized through multiple different methods(22, 23). Some of the genetic determinants of lipid metabolism in *R. opacus* are known: the TAG storage mediator *tadA*(24), and multiple wax ester synthase/DAG

O-acyltransferase (WS/DGATs)(25), further positioning *R. opacus* as a potentially effective chassis for bioconversion of lignin.

Towards this eventual goal *R. opacus* was been successfully adapted for increased growth and TAG production after ~270-420 generations on a range of MLBP mixtures(19). However, the sparse number of acquired single nucleotide polymorphisms in each adapted mutant was insufficient to determine a causative genetic mechanism of adapted tolerance. The highest performing mutant, PVHG6, was characterized against WT for differential expression (DE) of the transcriptome, but an observed lack of DE in the phenolic catabolism, degradation, and transport gene clusters after adaption precluded the identification of a regulatory based mechanism which might explain the increased tolerance to MLBPs as well.

We hypothesized that *R. opacus PD630* adapted to increasingly complex phenolic compound mixtures as an elementary facsimile of real-world lignin would result in strains with increasingly similar transcriptional profiles to *R. opacus* grown in real-world lignin. Moreover, characterization of more adapted strains could aid in identifying genetic elements universally important for tolerance to aromatic perturbation. In this study, we expand the phenotypic characterization of mutants adapted to MLBP mixtures, comparing DE not only to WT, but also each novel genotype at higher concentrations of each MLBP mix, when the adapted tolerance mechanism is at its strongest effect. We confirm that PV1 and PVHG6 exhibit increased growth versus WT at the same concentration, and that this is due to improved utilization of vanillic acid. These dynamics are concentration dependent; at high concentrations of MLBPs non-permissive to WT growth, all evolved strains exhibit increased utilization rates, but PVH5 and PVHG6 preferentially utilize 4-hydroxybenzoate before other carbon sources. We identify a divergent expression signature linking growth of PV1 and PVH5 at high concentrations of MLBPs to

increased expression of the β-ketoadipate pathway important for aromatic compound catabolism. Finally, we identify universally differentially expressed novel putative operon with an expression signature similar to the β-ketoadipate pathway which may play a role in tolerance to aromatic MLBPs.

5.2 Results

To contextualize compositional alterations to the *R. opacus PD630* transcriptome after adaptation to MLBP mixtures, we compared the substrate utilization, growth rates, and transcriptomic expression profiles of strains adapted to 2-4 phenolic compounds: *PV1* (Phenol + Vanillic acid), *PVH5* (Phenol + Vanillic acid + 4-hydroxybenzoic acid), and *PVHG6* (All preceding compounds + guiacol), to WT strains. We also included a condition growing each mutant strain at a concentration inhibitory for WT growth. This allowed us to observe adapted mutants in conditions in which evolved mechanisms for tolerance would be at maximum expression (see figure S1 for carbon source concentrations and glossary of terms).

5.2.1 Adapted mutants display concentration-dependent aromatic catabolite repression, increased tolerance, and consumption

We began by first measuring the ability of each mutant strain to grow and consume MLBP sources they were evolved on. We grew WT as well to determine a baseline utilization rate for *R. opacus* (figure 1a,d,g).

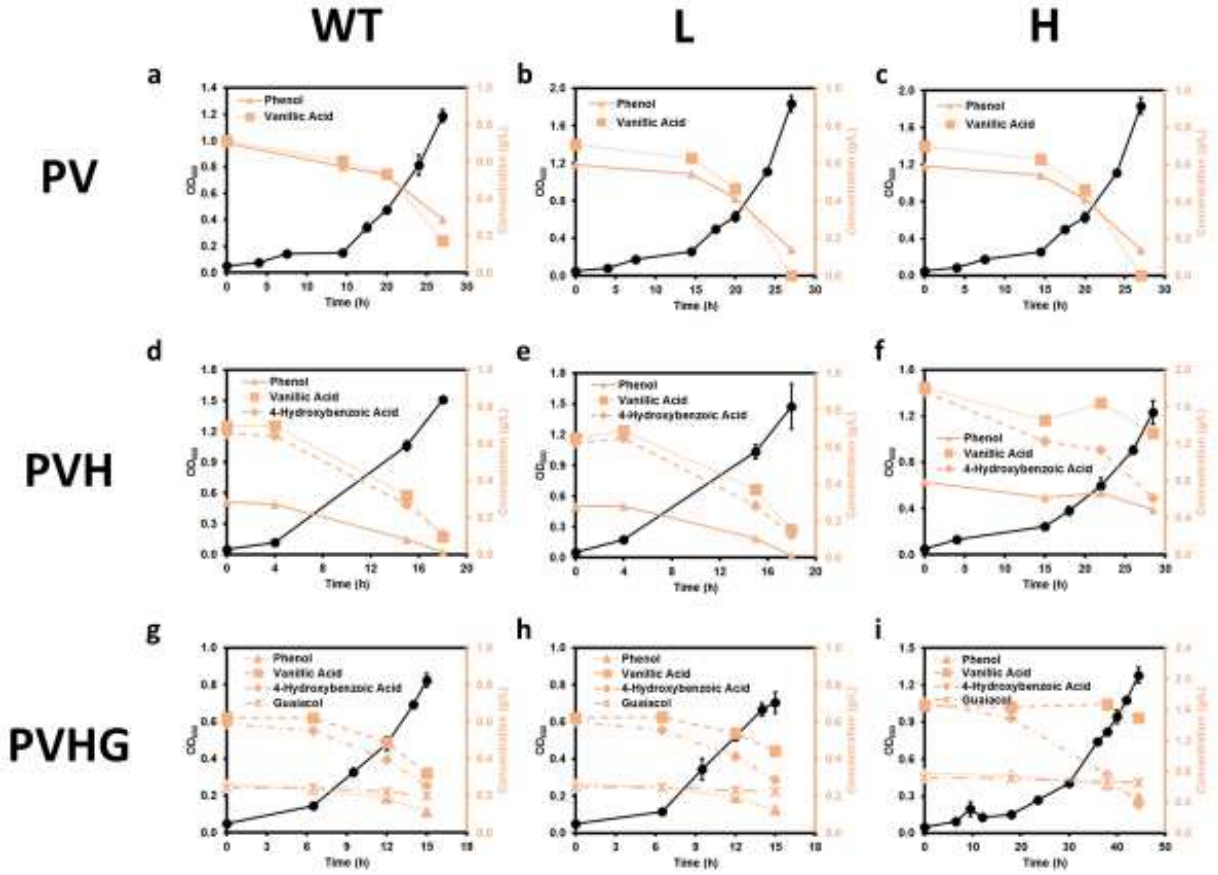


Figure 5. 1 Adapted mutants increase utilization of MLBPs. a-i. Rows designate the MLBP each strain (column) was grown in. The last two columns, L and H, are placeholder descriptions referring to the mutant at the WT "low" concentration and at the WT non-permissive "high" concentration of each MLBP. The left sided Y-axis refers to cell density as measured by optical density (OD_{600}). The right-sided Y axis refers to MLBP concentration available in the supernatant over time (g/L). Salmon colored lines indicate the concentration over time for each aromatic compound present in the MLBP, as indicated by the symbol legend. The x-axis is time measured in days.

WT co-utilized phenol and vanillic acid, utilizing more vanillic acid (figure 1a). Previous reports indicated *R. opacus* induces catabolite repression when exposed to 4-hydroxybenzoate and benzoate(19), preferentially degrading these compounds over phenol, vanillic acid, and guaiacol. Here we see no catabolite repression of phenol and vanillic acid when fed 4 hydroxybenzoate at the WT permissive concentration, with both WT grown in PVH and PVHG utilizing phenol, vanillic acid, and 4-hydroxybenzoate simultaneously (figure 1d,g). There was also no significant change in the concentration of guaiacol by 15 hours during WT growth in

PVHG, the only compound not significantly utilized by *R. opacus* (see Table 5.2S for all significance tests).

Adapted mutants grown at the WT permissive concentration of MLBPs exhibited enhanced utilization characteristics, with one exception: PVH5 (figure 1e) did not increase the utilization of any of the aromatic compounds and had no significant change in growth rate. PV1 (figure 5.1b) co-utilized phenol and vanillic acid but was able to increase the rate of vanillic acid consumption versus the ancestral strain (p value = 0.0047, Students t-test). PVHG6 also increased utilization of vanillic acid (p = 0.0318, Student's t-test, figure 5.1h), resulting in both PV1 and PVHG growing significantly faster than WT at the same concentration (p value = 0.0005 and p value = 0.040, respectively, Student's t-test).

Similar dynamics for utilization of MLBPs are observed at higher concentrations. PV1 (figure 5.1c) exhibited an increased growth rate (p = 0.001, Student's t-test), as well as a faster rate of vanillic acid utilization for (p = 0.0018, Student's t-test). Again, PVH5 at a higher concentration did not increase utilization of any of the MLBPs, though it did achieve a faster growth rate (p = 0.0065, Student's t-test, figure 5.1f). PVHG6-H also utilized MLBPs faster than PHVG-L (p = 0.001, Student's t-test, figure 5.1i), and increased the rate of utilization for phenol and 4-hydroxybenzoate (p = 0.038 and p = 0.001, Student's t-test). PVH5 and PVHG6 at low concentrations co-utilized phenol, vanillic acid, and 4-hydroxybenzoate, similar to WT, but when challenged with a higher concentration of MLBPs, preferentially utilized 4-hydroxybenzoate before phenol or vanillic acid. As seen previously with WT and PVHG6-L, PVHG6-H growth did not significant change guaiacol concentration in the media.

5.2.2 Adaptation changes the transcriptome of R. opacus pd630 grown in diverse carbon sources

We next conducted redundancy analysis (RDA) on the first two axes of a principal component analysis of the 500 most highly variable transcripts to test the effects of carbon source and strain on the PCA coordinates of all samples (figure 5.2a).

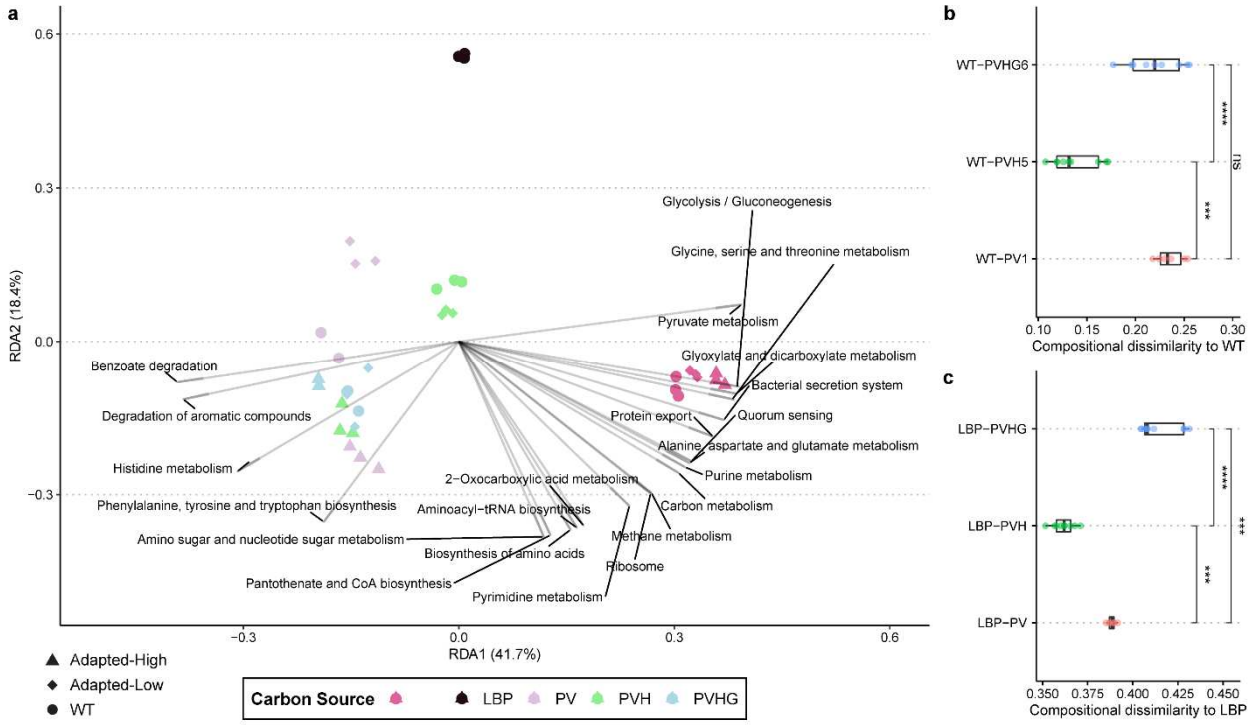


Figure 5. 2 ***Expression profiles of MLBP mutants do not cluster with WT grown in LBP. a.*** *PCA analysis of the 500 most highly variable loci in the transcriptome. Redundancy analysis was used to generate a constrained ordination. The best model using the summed abundance of R. opacus KEGG pathways, strain, and condition as terms was estimated using stepwise permutation (999). All remaining KEGG pathways used as terms were then fit to the ordination ($p < 0.05$, envit function with 999 permutations) and the vectors of greatest variation for each pathway are displayed. Circles denote WT, diamonds denote adapted strains-L, and triangles represent adapted strains-H, and are color coded by carbon source.* ***b****. Evolved strain centroid distance to WT. Strain and Carbon Source were used as terms in the redundancy analysis. Each boxplot represents the distance between the estimated centroids distance for each strain to the centroid generated from WT samples grown in the same MLBPs. PVHG5 is the most similar to WT.* ***c****. Carbon source centroid distance to LBP. WT grown in PVHG6 is the most dissimilar to WT grown in LBP.*

Both Strain (Goodness-of-fit = 0.8291 R2 and P < 0.000999, 999 permutations) and Carbon Source (Goodness-of-fit = 0.7098 R2 and P < 0.000999, 999 permutations) were highly explanatory. We next assessed the correlation between the compositional differences between all conditions and variation in summed KEGG pathway abundances by fitting linear trend surfaces of all annotated *R. opacus PD630* KEGG pathways, identifying 22 significantly well explained by the ordination (adjusted P < 0.05 and r > 0.6).

Many of the samples grown in the three MLBP mixtures varied along RDA2 and were close to fitted vectors for the benzoate and aromatic compound degradation pathways, as well as the phenylalanine, tyrosine, and tryptophan biosynthesis pathways. We also included comparator strain Glc1 (*R. opacus* adapted for increased growth on the simple sugar glucose) samples grown at Low and High concentrations, alongside WT grown in glucose. The fitted vectors for glycolysis/gluconeogenesis, glyoxylate and dicarboxylate metabolism, and glycine, serine, and threonine metabolism lay closest to samples from both the WT and adapted strain Glc1 grown in glucose. We next determined the chord distance between all adapted strains and the WT strain grown in the same MLBPs to determine whether increasing MLBP complexity resulted in greater expression dissimilarity compared to WT. Interestingly, we did not find a relationship between increasing MLBP complexity and compositional difference to WT after adaption, with PV1 and PVHG6 both producing a similarly high level of distance between the expression profiles of WT also grown in the same concentration (adjusted p > 0.05, wilcox test, figure 5.2b). When comparing WT grown in MLBPs to WT grown in real world LBP, WT grown in PVHG produced significantly higher compositional dissimilarity compared to WT grown in either PVH5 or PV1 (adjusted p = 0.0008 and adjusted p = 0001, wilcox test, figure 5.2c). In both cases the most complex MLBP, PVHG6, consistently produced large amounts of dissimilarity.

5.2.3 MLBP mutants display minimal evidence of a shared transcriptional state after adaption and divergent regulation of the β-ketoadiapate pathway at higher concentrations

After observing differences in the composition of expression profiles grown on different carbon sources, we next identified the shared highly differentially expressed genes between WT and the adapted strains grown in each MLBP mixture (defined as adjusted $P < 0.01$ and log2 fold change > 2, figure 5.6S). There were only 4 shared differentially expressed genes, the tetR/acrR transcriptional regulator K2Z90_RS06615, a glycosyltransferase family 2 protein (K2Z90_RS11395), an esterase (K2Z90_RS37430), and the agmatinase *speB* (K2Z90_RS37435). Though this still represents a significant higher sharedness than is expected by chance (P value = 3.33e-07, exact test of multi-set intersection), there was also no shared enrichment of any of the MLBP degradation clusters previously identified(19) (figure 5.7S).

To observe the effects of adaptation on the regulation of aromatic catabolism, we investigated differences in expression of the β-ketoadipate pathway across conditions (figure 5.3). There was very little DE in any of the three pathway clusters between the adapted strains and WT, and PVH5 was the only strain which upregulated any part of the b-ketoadipate pathway (4/5 genes from cluster 3, figure 5.7S). PV1, though it exhibited an increased growth rate versus WT, significantly downregulates 4/8 genes within cluster 1. When comparing DE between the adapted strain at the high and low concentrations, however, a divergent expression pattern emerged. While PV1 and PVH5 significantly increase the expression of 14/16 and 16/16 genes within the pathway, PVHG6 downregulates 6/16 genes, including 3 out of the 5 genes within pathway cluster 3.

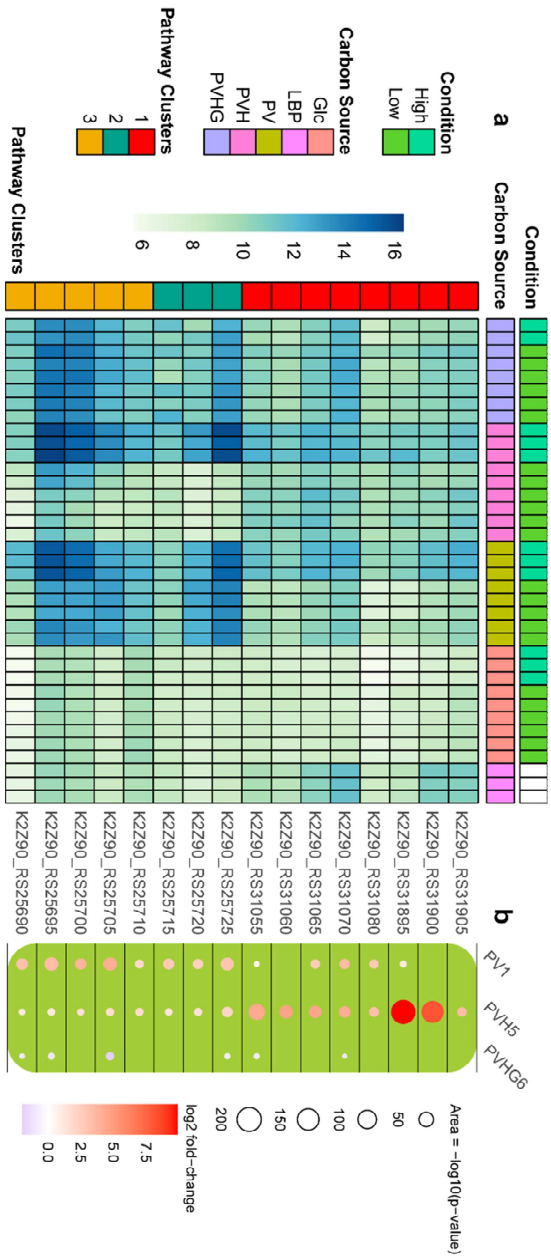


Figure 5. 3 Adapted strains display divergent expression of β-ketoadipate pathway. a. Heatmap of expression values for the β-ketoadipate pathway for all samples. Cell colors denote R-log transformed expression values. Column annotations identify Strain and Carbon Source for each sample (column). Row annotation bar identifies the pathway cluster each loci belongs to. Loci names are row labels shared between sub-plot a and b. b. Bubblegum plot of differentially expressed loci within the β-ketoadipate pathway. Each column contains circles representing the differential expression of one loci within a comparison between an adapted strain grown at high concentration versus the adapted strain grown at a low concentration. Size of the circle represents adjusted p value, and color represents Deseq2 log2foldchange values for each comparison.

When comparing the sets of highly differentially expressed genes between each phenolic MLBP mutant at low concentrations of MLBPs versus the higher concentration, we identified a set of 32 differentially expressed genes at the union of all sets, representing a 5.73-fold enrichment over expected (adjusted P = 3.362869e-13, exact test of multi-set intersection, figure 4.4). 6 of these genes were located in a similar genomic context. Expanding the analysis to all DE genes (adjusted $p < 0.05$) revealed a contiguous set of loci which were differentially expressed in a similar fashion to the β-ketoadipate pathway, with PV1 and PVH5 upregulating the set of genes, while PVHG6 exhibited downregulation (figure 5.4). of a putative 8-loci operon (K2Z90_RS17585 - K2Z90_RS17730) containing NAD(P)/FAD-dependent, SDR and PDR/VanB family oxidoreductases, aldehyde and acyl-CoA dehydrogenases, the putative long chain fatty acid-CoA ligase *fadD11* (see table 5.2S for the list of 6 universally DE genes). Immediately after this set of genes were a set of four loci which were removed from analysis due to extremely low transcript counts across all samples, and the loci preceding were not differentially expressed in PVHG6.

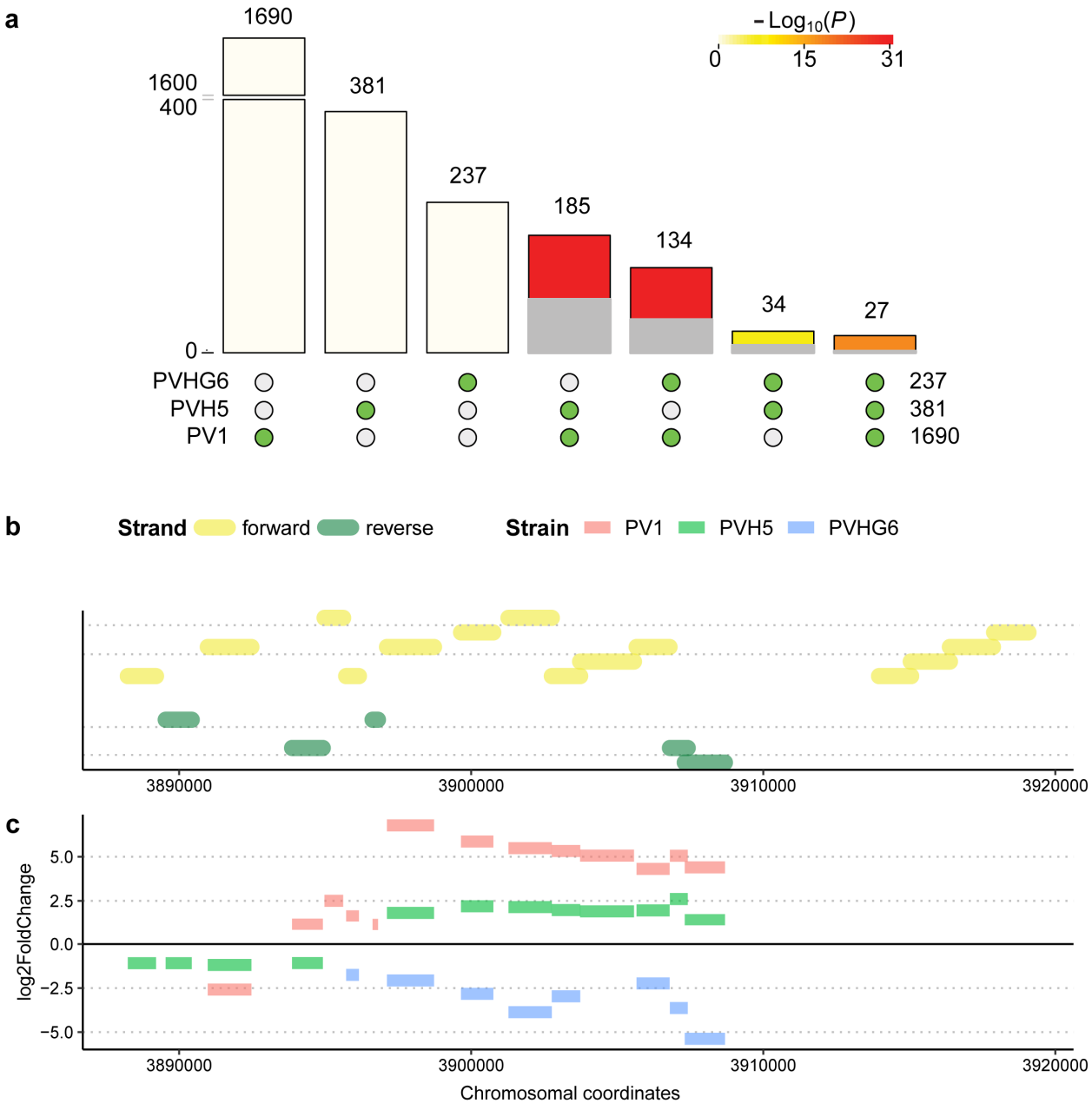


Figure 5. 4 Set analysis identifies loci divergently expressed between adapted MLBP mutants at high concentrations. a. Upset plot showing the shared highly differentially expressed genes (log2foldchange > 2 and adjusted Deseq2 p value < 0.01) between different groupings of the adapted MLBP strains. Color of the bar denotes the p value of the significant overrepresentation of that group within all sets versus the amount expected from random sampling, which is shaded in grey. Numbers floating above bars denote the set totals, and the dot plot below the barplot identifies the DE from each strain in the set. There were 27 loci differentially expressed in all three adapted MLBP mutants at high concentrations, 4 of which were located next to each other within a novel operon. b,c. Genomic context and fold-change values of loci within a novel operon DE among all MLBP mutants at high concentration. Chromosomal coordinates of each loci within the genomic context of the novel operon, with color denoting strand. In c

the log2foldchange (when significant, Deseq2 adjusted p < 0.05) of each loci is displayed, and color-coded for each strain. PV1 and PVH5, similar to the expression of the β-ketoadipate pathway (figure 3), significant upregulates a novel operon of genes while PVHG6 diverges from this expression signature.

5.2.4 PVHG6 exhibits higher transcriptomic resilience than mutants grown on less complex MLBPs

Finally, we compared the effect of increasing the concentration of each carbon source on the transcriptome of each adapted mutant. This was conducted by taking the chord distance between all mutant transcriptomes at low and high concentrations. We observed a non-linear relationship between increasing MLBP complexity and decreasing chord distance, with *PV1* having the greatest perturbation in expression profiles (and thus the least amount of transcriptomic robustness) between concentrations. PVHG6, on the other hand, exhibited the smallest change in expression profile during MLBP perturbation (figure S3). We observed that the change in dissimilarity with increasing carbon source complexity followed a power low, and when transformed to log space, this relationship became linear (R_{adj}=0.93, p = 0.001):

Equation 5.1

$$y = -0.161 - 0.961x$$

5.3 Materials and Methods

5.3.1 Supernatant compound concentration analysis:

Supernatant concentrations of chemicals were determined as described:

Phenol only

Phenol concentration was determined via UV-Vis spectroscopy by monitoring supernatant absorbance at 270 nm.

Glucose only

Glucose concentration in the supernatant was measured using an HK glucose kit purchased from Sigma Aldrich (part number GAHK20-1KT).

PV and PVH

Phenol, vanillic acid, and 4-hydroxybenzoate concentrations were measured using an Agilent 5977C GC-MS with a polysiloxane column, using a temperature ramp curve of 200 to 315 °C over 55 minutes.

PVHG

Supernatants were derivatized with methyl chloroformate (MCF) using a procedure outlined in the literature(26) and subsequently tested using GC-MS.

5.3.2 Strains and DNA manipulation

R. opacus PD630 (DSMZ 44193) was used as the WT strain. PV1, PVH5, PVHG6, and Glc1 were all generated from previous work(19).

5.3.2 Growth assay biomass measurement:

Carbohydrate extraction

10 mg of dry cells were hydrolyzed at 98 °C for 3 h using 3 mL of 2.5 N HCl, and then neutralized with solid sodium carbonate. The sample was diluted to 10 mL and centrifuged. The supernatant was then diluted 10:1 and 1 mL 5 wt% (w:v) phenol solution was added to the supernatant, followed by 5 mL of 96% sulfuric acid. The sample was shaken for 10 minutes and placed on a heating block at 30 °C for 20 more minutes. The color was measured at 490 nm, and total carbohydrate content was calculated using a calibration curve.

Lipid extraction

Lipid extraction, purification, and measurement was conducted using the Bligh and Dyer method(27).

Protein analysis

5 mg of frozen dry cells were sent to the University of California-Davis for amino acid concentration measurement. Cell samples were hydrolyzed using 200 µL of 6 N HCl and 1% phenol at 110 °C for 24 h. The hydrolyzed sample was dried and redissolved in NorLeu dilution buffer to 20 mL. The redissolved sample was vortexed, spun down, and 50 µL was injected into an L-8800 AAA Hitachi High-Speed Amino Acid Analyzer.

5.3.3 *R. opacus PD630* culture conditions:

Strains of Rhodococcus opacus PD630 were grown at 30°C and centrifuged at 250 rpm. Seed cultures from single isolates were grown in 0.2 g/L each phenol, vanillic acid, 4-hydroxybenzoate, guaiacol, and benzoate with 1 g/L (NH4)2SO4 as a nitrogen source. Growth assays were conducted by adding 45 mL of Greasham media was to 5 mL of each MLBP or glucose solution, and balancing pH to ~7.2 with 2 M NaOH and 1 N HCl. Wet cell weight was determined using cell culture optical absorbance at 600 nm (OD600).

3.3.4 RNA extraction and rRNA depletion:

RNA extraction proceeded as previously described(19). Briefly, RNA was extracted using the RNA MiniPrep kit (Zymo Research) and treated with two doses of TURBO DNase I (Ambion) for 30 minutes at 37C to remove DNA contamination. DNase-treated RNA was then cleaned using the RNA Clean & Concentrator Kit (Zymo Research), and then tested for DNA contamination by PCR amplification using intergenic primers. Any samples with distinct bands

after PCR were digested and cleaned again until no DNA was detected. Total RNA concentration was quantified using a NanoVue Plus spectrophotometer and rRNA was depleted using the Bacterial Ribo-Zero rRNA removal kit (Illumina). mRNA was converted to cDNA and barcoded using previously described methods(17). cDNA samples were then pooled in equimolar ratios, diluted in nuclease-free water to a final concentration of 10nM for sequencing.

5.3.5 Sequencing Library Preparation and Transcriptomic Analysis:

A 20 uL equimolar mix of cDNA samples were submitted for sequencing at the Center for Genome Sciences and Systems Biology in Washington University in St. Louis School of Medicine. Samples were single-end sequenced (1x75bp) using the Illumina Hi-Seq 2500 System.

After demultiplexing, raw reads were trimmed using trimmomatic(28) using the standard settings and the CROP length of 75bp. Samples with more than 15 million trimmed reads were subsampled using seqtk(29), and then mapped to a bowtie2 library built off the ASM2054278v1 *R. opacus* reference (Refseq assembly GCF_020542785.1)(30), sorted and indexed. Expression counts were calculated for each genetic loci using featureCounts(31) and imported into R(32) for statistical analysis and visualization. Differential gene expression analysis was performed using DESeq2(33). Redundancy analysis was performed using vegan(34) and biodiversityR(35). Heatmaps were generated in R using pheatmap(36), and set enrichment and plotting was conducted using SuperExactTest(37) in R. Visualizations were created using ggplot2(38).

5.4 Discussion

Adapting microbial chassis for increased tolerance to toxic substrate is a longstanding goal towards lignin valorization(16, 17, 19). The heterogeneity of lignin structure over time and plant stock, along with the intensive pre-treatment procedures, prohibited us from directly utilizing

lignin for bacterial adaption. Adapting *R. opacus* to MLBPs allows us insight into the mechanisms behind aromatic tolerance, and the effect of adaptation on its ability to exhibit resilience to further environmental perturbation. Given that each mutant was adapted over a similar timeframe, this further provides an opportunity to observe the effect of carbon compound complexity on the ability of *R. opacus* to adapt.

Adaption implies improvement, and in this case, we observe an increase in growth rate and utilization over WT for PV1 and PVHG6. Interestingly, PVH5 does not increase growth rate or utilization of phenol, vanillic acid, or 4-hydroxybenzoate, even though PVH5 is the only strain which upregulates the β-ketoadipate pathway (cluster 3) versus WT. Cluster three is involved in catabolism of phenol, versus cluster 1 which catabolizes vanillic acid and 4-hydroxybenzoate. PVH5 thus adapted to increased concentrations of phenol by upregulating the phenolic catabolism cluster of the β-ketoadipate pathway. This is in sharp contrast to PV1 and PVH5, which significantly increase the utilization rate of vanillic acid over WT. Recent work confirmed that *R. opacus* converts about 92%(mol/mol) of the conjugate base vanillate to muconic acid via the β-ketoadipate pathway(39), and further work is needed to determine whether the increases in utilization rate of vanillic acid result in increased conversion rates.

At higher concentrations that PVH5 and PVHG6 preferentially utilize 4-hydroxybenzoate, a characteristic not detected at lower concentrations. Catabolite repression in *Rhodococcus opacus* has been observed previously for benzoate(40), phthalate(40), 4-methoxybenzoate(41), and WT and PVHG6 both exhibited 4-hydroxybenzoic acid preference when grown at 2.5 g/L total aromatics(19). This is a substantially higher concentration of total aromatics than the WT permissive concentrations for PVH and PVHG in this study (1.5 g/L for both, see Figure 5.5S), but less than the high concentration (3.75 and 4.5 g/L respectively). Evolutionary theory suggests

that catabolite repression occurs when a bacterium faces a resource allocation trade-off between singular and multiple enzyme production(42). We see no evidence of this in the expression of the degradation clusters for each aromatic carbon source, and the entire β-ketoadipate pathway was upregulated in PVH5, which would indicate catabolite repression is occurring outside expressional regulation.

We expected that after adaptation to extreme concentrations of MLBP's, growing the adapted strain at a lower, WT-permissive concentration would result in minimal perturbation between WT and mutant expression profiles. This was observed for PVH5, which produced a transcriptional state at low concentrations which was compositionally more similar to the ancestral strain than either PV1 or PVHG6, with fewer than 100 highly differentially expressed genes compared to WT grown in PVH. PVHG6 samples grown at the high concentration was more dissimilar to the ancestral strain grown in PVHG as well as in lignin. This suggests that the method of tolerance for the new transcriptome state of PVHG6 after adaptation is permanent, resulting in constitutive activation of genetic tolerance elements even at a low concentration. We observe this again when comparing the resilience of the adapted state to increased concentration: the PVHG6 strain produced the least amount of compositional dissimilarity. This relationship between the increase in MLBP complexity and difference in compositional profile at a higher concentration in adapted mutants followed a power law(43). We can conclude that the changes which occur during adaptation to increasingly complex carbon sources result in entrenchment of alterations to gene expression. Thus, the mutant becomes increasingly less able to return to an ancestral state, presumably as more alterations to metabolism and genetic regulation occur.

As expected, we observed that adapted mutants PVH5 and PV1 at high concentrations correlated with increases in the aromatic and benzoate degradation pathways. We further detected

upregulation of the b-ketoadipate pathway in PV1 and PVH5 mutants at high concentrations; direct confirmation that tolerance to high concentrations after adaption results in mutants in "tunable" expression of the aromatic utilization pathway(44). PVHG6 grown in both low and high concentrations of the most complex MLBP does not exhibit this behavior. Instead, it diverges from the expression profiles of *PV1* and *PVH5* and is unable to increase expression of the b-ketoadipate pathway at high concentrations. This could be evidence for PVHG6 utilizing a different method of adaption to high concentrations; alternatively, it is also possible that the high PVHG concentration was closer to the minimum inhibitory concentration as it was a higher concentration of total aromatics than either PV-H or PVH-H. Regardless, this is confirmation of a subtle interplay between the straight-forward relationship between carbon mixture complexity and resilience after adaption, and evidence of divergent expression after adaption.

Even more surprisingly, WT grown on LBP exhibits limited expression of pathway clusters 2 or 3. These pathways make up the catechol branch of the β-ketoadipate pathway, and are responsible for the utilization of phenol, benzoate, and guaiacol. This suggests that WT on LBP is not utilizing phenol, either due to its absence or the presence of a less recalcitrant food source. Or perhaps there are other aromatic catabolism pathways within the *R. opacus* genome which are not known. There is already some evidence of this: Cai et al. observed only a 0.10%±0.02 g/g *cis-cis* muconate yield when *R. opacus* engineered without a fully functional β-ketoadipate pathway was grown on vanillin, and they attributed this to unknown auxiliary utilization pathways(21). We observe divergent expression in a novel energy operon for all three MLBP mutants, suggesting it may be directly related to adaptation to aromatic carbon sources. Genes with similar annotations have been identified and described in *Rhodococcus*: 2 aromatic hydrolases specific to the meta-cleavage of ethylbenzene have been identified in *R. opacus RHAI*

and a PDR/vanB family oxidoreductase is responsible for the conversion of vanillin to catechol(45). It is puzzling that PVHG6 downregulates the β-ketoadipate pathway as well as this novel operon, and further work is needed in order to elucidate the exact mechanisms by which these different adaptive strategies result in increased tolerance.

Information on the adaptive trajectories of *R. opacus PD630* after adaptation to extreme environments is generally informative, but a definition of the adapted transcriptional state when grown in real world LBP is needed to contextualize these findings. Due to experimental complexity, we were unable to grow each adapted mutant in lignin, or to generate a mutant adapted on real world lignin. This meant we were unable to directly test whether adaption to increasingly complex MLBPs provides added tolerance in PVHG6 versus PV1. Instead, we observed that the WT *R. opacus* transcriptional state in complex MLBP mixtures does not result in increased compositional similarity to *R. opacus PD630* grown in real world lignin. Indeed, we find no clear relationship between model compound mixture complexity and compositional similarity to WT-LBP expression profiles. Aromatic compound degradation clusters and the b-ketoadipate pathway, both necessary for growth on MLBPs, are largely not expressed at similar levels when grown on LBP. This is surprising, given that each of the phenolic compounds used in the adaptation study: phenol, vanillic acid, 4-hydroxybenzoate, and guaicol, can be present in real-world lignin breakdown products, along with a myriad range of ketones, alcohols, esters, and other aromatics. These data call into question the validity of the proposed model that adaptation to increasingly complex lignin-like carbon sources will yield mutants with increased utilization on real-world lignin(19). There was no apparent relationship between MLBP complexity and similarity to the WT-LBP transcriptome. Further work is needed to integrate this

observation into a coherent understanding of a relationship between tolerance and carbon source complexity in *R. opacus*.

In this study, we compare WT expression profiles in MLBPs to WT grown in LBP, as well as to each adapted strain, to better contextualize the different transcriptional states required for MBLP versus LBP utilization and the adaption trajectories of each mutant. We identify increased utilization of vanillic acid as important for the increased growth of adapted *R. opacus*, but not all adapted strains. This heterogeneity in utilization is seen repeatedly in the analysis of the transcriptomic profiles of adapted strains grown on MLBPs. Along with a lack of a consistent relationship between increasing MLBP complexity and similarity to WT grown in MLBPs or real-world lignin, it is likely that multiple strategies for aromatic tolerance have been adapted. In multiple strains, the ability to survive at high concentrations of aromatic monomers is dependent on increased expression of the β-ketoadipate pathway but may also rely on the activation of novel genetic loci, a subject for future work. In this study we demonstrate the utility of top-down screens and identify shared mechanisms in *R. opacus* for tolerance and adaption to aromatics.

5.5 Acknowledgements

This work was co-authored by Winston Anthony, Rhiannon Carr, Zeyu Shang, Aditya Ponukumati, Bin Wang, Jie Ning,Yingjie Tang, Marcus Foston, Tae-Seok Moon, and Gautam Dantas, *In preparation*. Funding from U.S. Department of Energy [DE-SC0018324 to M.F., G.D., and T.S.M.]. All authors contributed text to the manuscript. W.E.A., R.C., and A. P. generated the figures. W.E.A, R.C., Z.S., Y.T., M.F., T.S.M., and G.D. designed experiments. R.C., Z.S., and W.E.A ran experiments and collected data. The authors thank members of the Moon, Dantas, and Foston research groups for scientific discussions.

5.5 Appendix (Supplemental Material)

5.5.1 Tables

Table 5. 1S Loci differentially expressed by all MLBP mutants at high concentrations

Annotation	Locus Tag
metal-dependent hydrolase	K2Z90_RS17595
PDR/VanB family oxidoreductase	K2Z90_RS17600
aldehyde dehydrogenase family protein	K2Z90_RS17610
SDR family oxidoreductase	K2Z90_RS17615
acyl-CoA dehydrogenase family protein	K2Z90_RS17625
hypothetical protein	K2Z90_RS17630

5.5.2 Figures

Experiments

Substrate	Strain	Initial Concentrations (g/L)					
		P	V	H	G	B	GLC
P	WT	0.75					
	P1-L	0.75					
	P1-H	1.55					
PV	WT	0.70	0.70				
	PV1-L	0.70	0.70				
	PV1-H	1.35	1.35				
PVH	WT	0.30	0.60	0.60			
	PVH5-L	0.30	0.60	0.60			
	PVH5-H	0.75	1.50	1.50			
PVHG	WT	0.25	0.50	0.50	0.25		
	PVHG6-L	0.25	0.50	0.50	0.25		
	PVHG6-H	0.75	1.50	1.50	0.75		
B	WT					6.00	
	B2-L					6.00	
	B2-H					14.10	
GLC	WT						250.00
	GLC1-L						250.00
	GLC1-H						300.00

P	Phenol
V	Vanillic acid
H	4-hydroxybenzoic acid
G	Guaiacol
B	Benzoate
GLC	Glucose
P1	P-adapted *Rhodococcus*
PV1	PV-adapted *Rhodococcus*
PVH5	PVH-adapted *Rhodococcus*
PVHG6	PVHG-adapted *Rhodococcus*
B2	Benzoate-adapted *Rhodococcus*
GLC1	Glucose-adapted *Rhodococcus*
WT	Wildtype *Rhodococcus*

Figure 5. 5S Experimental Design and glossary of terms. Graphical Layout of the carbon source mixtures and concentrations used for all growth and consumption experiments. Measurements are in g/L of total aromatics (or in the case of glucose, sugars) in the media at the introduction of R. opacus culture. WT and mutant-L conditions were grown at the same concentration, while the mutant-H conditions were grown at a higher concentration of total aromatics non-permissive to WT growth. The glossary identifies the letter symbols used for referring to the different combinations of aromatic carbon mixtures.

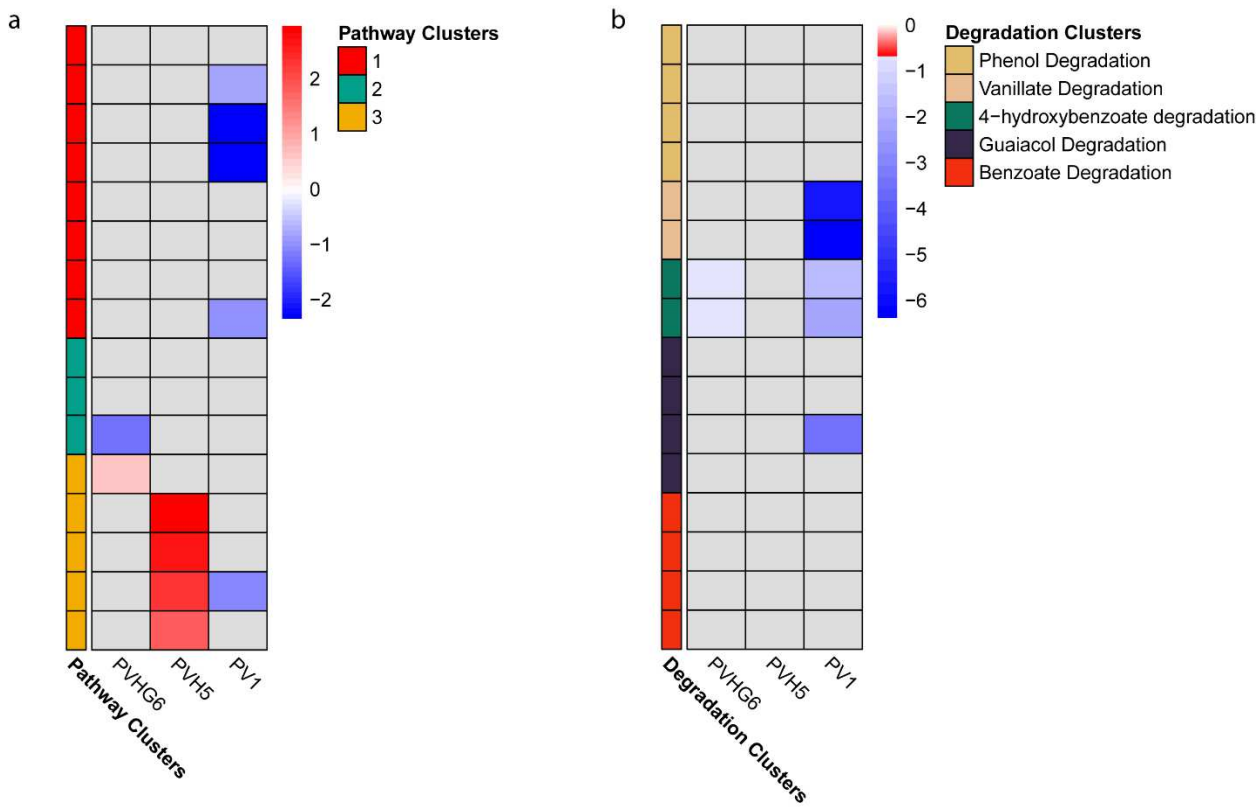


Figure 5. 6S Differential expression of the β-ketoadipate pathway and aromatic degradation clusters in evolved mutants versus WT. a. Heatmap of the differentially expressed loci in the β-ketoadipate pathway. Columns represent log2foldchange values from Deseq2 for MLBP evolved mutants compared to WT R. opacus grown at the same mixture concentration. Row annotations denote the different pathway clusters each loci belong to. Grey cells are loci which were not significantly expressed (Deseq2 adjusted p value < 0.05), while blue and red represent under- and over- expression compared to WT. PVH5 is the only strain which overexpresses genes within the β-ketoadipate pathway, overexpressing a majority of cluster 3. b. Heatmap of differential expression of the aromatic degradation clusters in MLBP mutants versus WT grown at the same concentration. Grey, blue, and red cells represent the same values as in a. Row annotations denote the aromatic carbon source degradation cluster each loci is associated with. There are no upregulated genes in this cluster for any of the mutant strains.

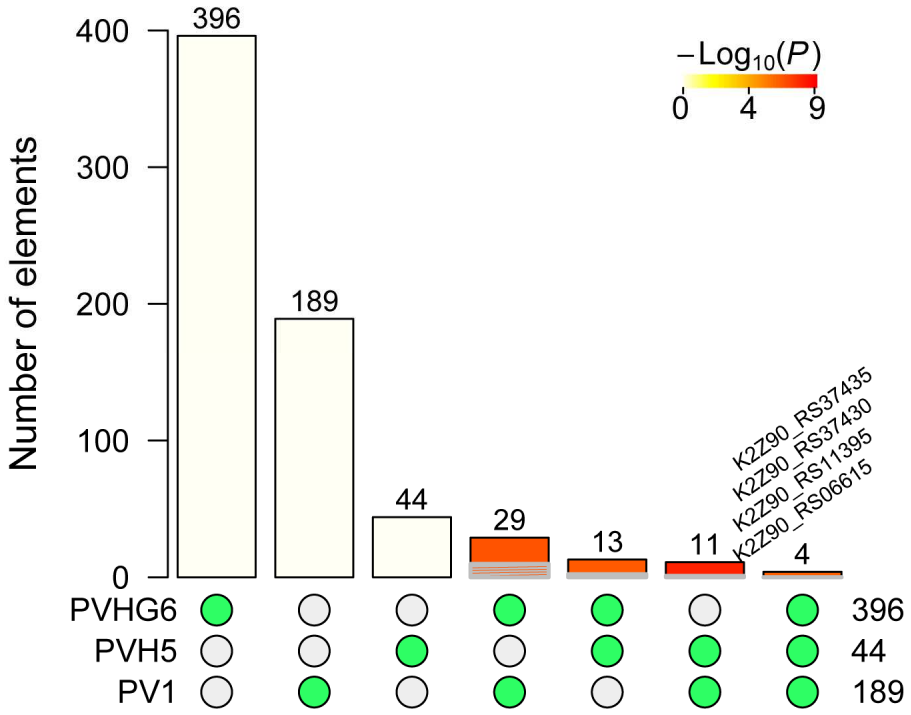


Figure 5. 7 Set analysis identifies loci divergently expressed between adapted MLBP mutants at low concentrations. Upset plot showing the shared highly differentially expressed genes (log2foldchange > 2 and adjusted Deseq2 p value < 0.01) between different groupings of the adapted MLBP strains. Color of the bar denotes the p value of the significant overrepresentation of that group within all sets versus the amount expected from random sampling, which is shaded in grey. Numbers floating above bars denote the set totals, and the dot plot below the barplot identifies the DE from each strain in the set. There were 4 loci differentially expressed in all three adapted MLBP mutants at low concentrations.

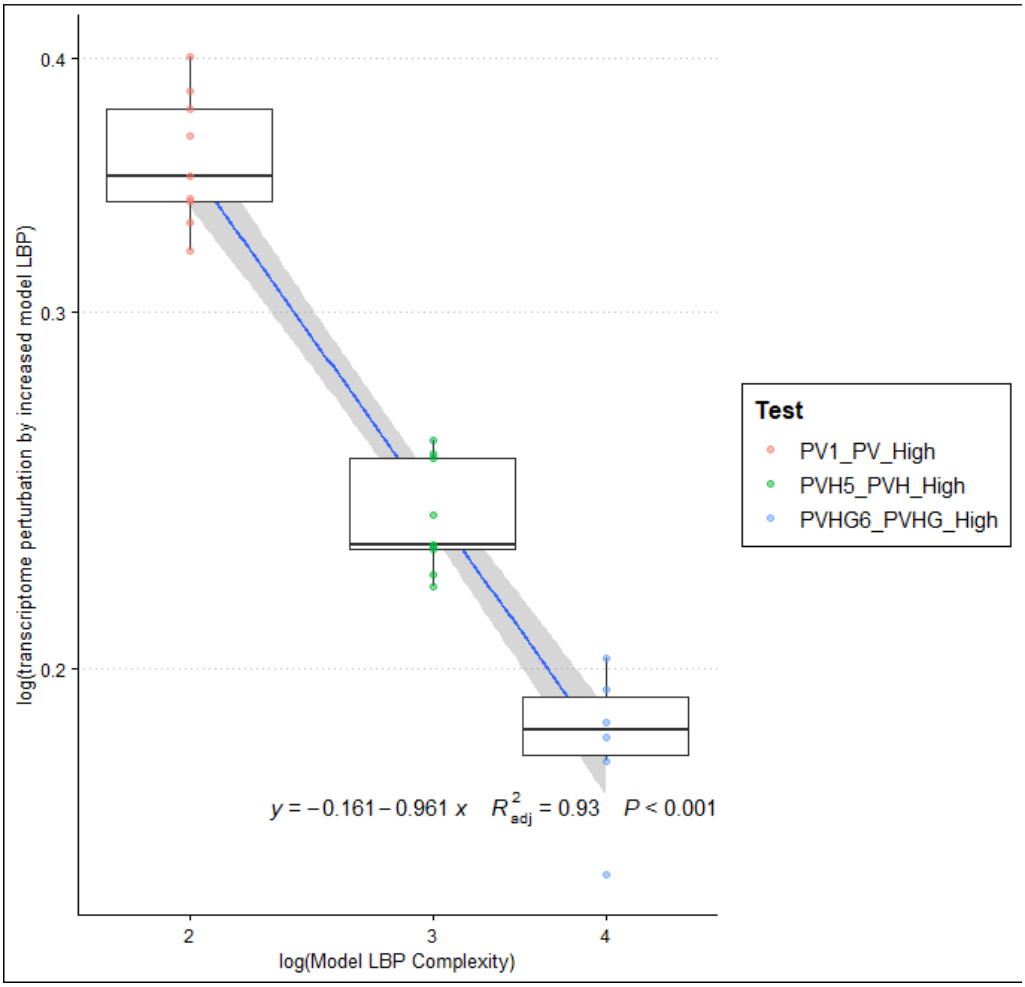


Figure 5. 8S PVHG6 exhibits transcriptomic resiliency when challenged with increased concentrations of MLBPs.

5.5 References

1. US Energy Information Administration Monthly Energy Review. August; 2022.

2. State TUSDo, President atUSEOot. The Long-Term Strategy of the United States: Pathways

to Net-Zero Greenhouse Gas Emissions by 2050. . 2021.

3. Agriculture USDo. Agricultural Innovation Agenda: Scoreboard Report. 2020.

4. Harkin JM. LIGNIN AND ITS USES. FOREST PRODUCTS LAB MADISON WIS; 1969.

5. Saha BC, Jordan DB, Bothast RJ. Enzymes, Industrial (overview). In: Schaechter M, editor. Encyclopedia of Microbiology (Third Edition). Oxford: Academic Press; 2009. p. 281-94.

6. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. Lignin Biosynthesis and Structure. Plant Physiology. 2010;153(3):895-905.

7. Giordano A, Liu Z, Panter SN, Dimech AM, Shang Y, Wijesinghe H, et al. Reduced lignin content and altered lignin composition in the warm season forage grass Paspalum dilatatum by down-regulation of a Cinnamoyl CoA reductase gene. Transgenic Res. 2014;23(3):503-17.

8. Patton AR, Gieseker L. Seasonal Changes in the Lignin and Cellulose Content of Some Montana Grasses. Journal of Animal Science. 1942;1(1):22-6.

9. Holladay JE, White JF, Bozell JJ, Johnson D. Top value-added chemicals from biomass-Volume II—Results of screening for potential candidates from biorefinery lignin. Pacific Northwest National Lab.(PNNL), Richland, WA (United States); 2007.

10. Mottiar Y, Vanholme R, Boerjan W, Ralph J, Mansfield SD. Designer lignins: harnessing the plasticity of lignification. Current Opinion in Biotechnology. 2016;37:190-200.

11. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, et al. Lignin valorization: improving lignin processing in the biorefinery. Science. 2014;344(6185):1246843.

12. Zakzeski J, Bruijnincx PCA, Jongerius AL, Weckhuysen BM. The Catalytic Valorization of Lignin for the Production of Renewable Chemicals. Chemical Reviews. 2010;110(6):3552-99.

13. Chio C, Sain M, Qin W. Lignin utilization: A review of lignin depolymerization from various aspects. Renewable and Sustainable Energy Reviews. 2019;107:232-49.

14. Beckham GT, Johnson CW, Karp EM, Salvachúa D, Vardon DR. Opportunities and challenges in biological lignin valorization. Current Opinion in Biotechnology. 2016;42:40-53.

15. Holder JW, Ulrich JC, DeBono AC, Godfrey PA, Desjardins CA, Zucker J, et al. Comparative and Functional Genomics of Rhodococcus opacus PD630 for Biofuels Development. PLOS Genetics. 2011;7(9):e1002219.

16. Kurosawa K, Laser J, Sinskey AJ. Tolerance and adaptive evolution of triacylglycerol-producing Rhodococcus opacus to lignocellulose-derived inhibitors. Biotechnology for Biofuels. 2015;8(1):76.

17. Yoneda A, Henson WR, Goldner NK, Park KJ, Forsberg KJ, Kim SJ, et al. Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. Nucleic Acids Res. 2016;44(5):2240-54.

18. Jiang W, Gao H, Sun J, Yang X, Jiang Y, Zhang W, et al. Current status, challenges and prospects for lignin valorization by using Rhodococcus sp. Biotechnology Advances. 2022;60:108004.

19. Henson WR, Campbell T, DeLorenzo DM, Gao Y, Berla B, Kim SJ, et al. Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. Metabolic engineering. 2018;49:69-83.

20. Mycroft Z, Gomis M, Mines P, Law P, Bugg TDH. Biocatalytic conversion of lignin to aromatic dicarboxylic acids in Rhodococcus jostii RHA1 by re-routing aromatic degradation pathways. Green Chemistry. 2015;17:4974-9.

21. Cai C, Xu Z, Xu M, Cai M, Jin M. Development of a Rhodococcus opacus Cell Factory for Valorizing Lignin to Muconate. ACS Sustainable Chemistry & Engineering. 2020;8(4):2016-31.

22. He Y, Li X, Ben H, Xue X, Yang B. Lipid Production from Dilute Alkali Corn Stover Lignin by Rhodococcus Strains. ACS Sustainable Chemistry & Engineering. 2017;5(3):2302-11.

23. Zhao C, Xie S, Pu Y, Zhang R, Huang F, Ragauskas AJ, et al. Synergistic enzymatic and microbial lignin conversion. Green Chemistry. 2016;18(5):1306-12.

24. MacEachran DP, Prophete ME, Sinskey AJ. The Rhodococcus opacus PD630 Heparin-Binding Hemagglutinin Homolog TadA Mediates Lipid Body Formation. Applied and Environmental Microbiology. 2010;76(21):7217-25.

25. Hernandez MA, Arabolaza A, Rodriguez E, Gramajo H, Alvarez HM. The atf2 gene is involved in triacylglycerol biosynthesis and accumulation in the oleaginous Rhodococcus opacus PD630. Appl Microbiol Biotechnol. 2013;97(5):2119-30.

26. Madsen RB, Jensen MM, Mørup AJ, Houlberg K, Christensen PS, Klemmer M, et al. Using design of experiments to optimize derivatization with methyl chloroformate for quantitative analysis of the aqueous phase from hydrothermal liquefaction of biomass. Analytical and Bioanalytical Chemistry. 2016;408(8):2171-83.

27. Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. Can J Biochem Physiol. 1959;37(8):911-7.

28. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.

29. Li H. Toolkit for processing sequences in FASTA/Q formats: lh3/seqtk. Github. 2022.

30. Firrincieli A, Grigoriev B, Dostálová H, Cappelletti M. The Complete Genome Sequence and Structure of the Oleaginous Rhodococcus opacus Strain PD630 Through Nanopore Technology. Front Bioeng Biotechnol. 2021;9:810571.

31. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013;41(10):e108.

32. Wickham H, Grolemund G. R for data science: import, tidy, transform, visualize, and model data: " O'Reilly Media, Inc."; 2016.

33. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550-.

34. Dixon P. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science. 2003;14(6):927-30.

35. Kindt R, Kindt MR. Package 'BiodiversityR'. Package for community ecology and suitability analysis. 2019;2:11-2.

36. Kolde R, Kolde MR. Package 'pheatmap'. R Package. 2018;1.

37. Wang M, Zhao Y, Zhang B. Efficient Test and Visualization of Multi-Set Intersections. Scientific reports. 2015;5(1):16923.

38. Wickham H, Chang W, Wickham MH. Package 'ggplot2'. Create elegant data visualisations using the grammar of graphics Version. 2016;2(1):1-189.

39. Zhou H, Xu Z, Cai C, Li J, Jin M. Deciphering the metabolic distribution of vanillin in Rhodococcus opacus during lignin valorization. Bioresource Technology. 2022;347:126348.

40. Choi KY, Zylstra GJ, Kim E. Benzoate Catabolite Repression of the Phthalate Degradation Pathway in Rhodococcus sp. Strain DK17. Applied and Environmental Microbiology. 2007;73(4):1370-4.

41. Karlson U, Dwyer DF, Hooper SW, Moore ER, Timmis KN, Eltis LD. Two independently regulated cytochromes P-450 in a Rhodococcus rhodochrous strain that degrades 2-ethoxyphenol and 4-methoxybenzoate. Journal of Bacteriology. 1993;175(5):1467-74.

42. Wang X, Xia K, Yang X, Tang C. Growth strategy of microbes on mixed carbon sources. Nature Communications. 2019;10(1):1279.

43. Schuster P. Power laws in biology: Between fundamental regularities and useful interpolation rules. Complexity. 2011;16(3):6-9.

44. Li N, Zeng W, Xu S, Zhou J. Toward fine-tuned metabolic networks in industrial microorganisms. Synthetic and Systems Biotechnology. 2020;5(2):81-91.

45. Yamada A, Kishi H, Sugiyama K, Hatta T, Nakamura K, Masai E, et al. Two nearly identical aromatic compound hydrolase genes in a strong polychlorinated biphenyl degrader, Rhodococcus sp. strain RHA1. Appl Environ Microbiol. 1998;64(6):2006-12.

Chapter 6:

Conclusions

"Micro"-biology implies diminutive, which is a misnomer. The multiple scales of organization and structure that produce life and community for bacteria are as important as they are in the largest organisms. In each chapter of this dissertation, I induce some change in microbes (either through their environment or genome) and using muti-omic techniques describe the methods bacteria use to recover or adapt.

The gut microbiome, beyond some initial characterization of population structure, is usually studied through the lens of illness and dysbiosis. During the past decade we have witnessed an explosion of literature attempting to link gut dysbiosis to other, more characterized diseases, such as Alzheimer's or cancers of the gastrointestinal tract. To me, this smacks of "the cart before the horse". How can we link dysbiosis to disease when the factors that pre-dispose a microbiome to dysbiosis are not well known? We know that some antimicrobials can cause dysbiosis, but is this a universal trait of pharmaceuticals? Similarly, can we identify individualized effects of antimicrobial perturbation and use that to detect individuals at greater risk for dysbiosis?

In Chapter 2 of my dissertation, I use metagenomic sequence analysis of experimentally perturbed microbiomes to answer these questions. It was of critical importance during the study development that we selected healthy volunteers with no antibiotic treatment in the last 6 months, as there is evidence that frequent repeated exposure to antimicrobials can produce

individualized changes and incomplete recovery(1). This resulted in 20 volunteers producing 289 stool samples before, during, and after treatment with 4 different antibiotics which were processed into whole shotgun metagenome sequences and for a subset, semi-quantitative culture for estimation of fecundity(2). After antibiotic treatment, there were universal signs of decreased species richness and colony forming units, which are consistent with previous reports identifying perturbation by antibiotics in healthy microbiomes(1, 3). However, we also identified antibiotic specific effects on the composition of volunteer microbiomes, with treatments producing a selection effect resulting in greater intertreatment dispersion between volunteers given Cefpodoxime (CPD) and Cefpodoxime with Azithromycin (CPD+AZM) and volunteers given Levofloxacin (LVX) or Azithromycin (AZM) alone. This has one primary consequence: An acute drop in species richness and enrichment of taxa able to escape the effects of treatment. Cefpodoxime enriches for the *Bacteriodetes* phylum, of which many species are intrinsically resistant to β-lactams(4). *Critically though, we observe no effects of these groupings on recovery time*. The process by which species re-colonize the microbiome after antibiotic perturbation is not related to the re-organization that occurs acutely, and analysis of the compositional variance between these groups on the last day of the study (185) confirmed no dissimilarity such as what was seen immediately after antibiotic treatment. Given these data, bacterial recolonization is likely affected instead by the environment and diet of the individual. Previous experimental work in animal models observed similar effects of environment and diet on the recovery of human microbiota, as well as altered recovery processes based on dietary fiber supplementation(5). Future work experimentally validating this last observation in human volunteers could lead to the development of antibiotic co-therapies to less acute effects on the gut microbiome.

Recovery was instead most strongly affected by a different treatment grouping: volunteers given azithromycin were delayed in recovery of species richness. This ~5 day delay is similar in length to the bioavailability of azithromycin, and is markedly longer than any of the other treatments(6-8). While azithromycin did not prevent volunteers from attaining a similar average number of species, it significantly increased *the time it took, and the compositional distance traveled to* get there. This means that azithromycin treatment results in volunteer gut microbiomes inhabiting these "intermediate states" for longer. The effects of this observation on overall human health are unclear, but it is possible that this could result in a longer period of instability during recovery; essentially keeping the door open for invasion by pathogens or pathobionts longer.

When microbiomes recovered, I observed strong evidence of community re-organization resulting in functional enrichment of antibiotic resistance in most, but not all, treatments. Levofloxacin did not significantly increase resistance burden, and this is most likely due to the most common form of resistance being mediated by single nucleotide polymorphisms instead of genes(9). Two resistance genes that increased significantly in the other treatments encoded resistance to tetracycline, which was not administered in this study. Taken together with the evidence of acute selection and community reorganization, this points to an unsettling conclusion: after antibiotic treatment the community reorganization that occurs is largely favorable to the entrenchment of broadly antibiotic resistant bacteria.

Finally, by connecting a volunteer's stool samples through principal component space, we were able to identify three volunteer microbiomes which underwent drastic changes in composition and diversity, eventually ending the study compositionally similar to microbiomes from critically ill ICU patients. These volunteers came from different treatments, which confirms

that while we do observe treatment-specific effects, there is no evidence for increased risk of microbiome dysbiosis by treatment. Instead, this points to individualized effects of antibiotic treatment on the dynamics of microbiome recovery, and we find support within the literature for this deduction (1, 10). Lloyd-Price et. al proposes modeling these dynamics as the movement of a ball across a stability landscape, where areas of health are represented as local minima(11). The ball (microbiome) normally rests within a healthy minima, and day to day variations in taxonomic and functional composition are minimal enough to keep the ball from being displaced. Antibiotic perturbation pushes the ball up out of the healthy minima, at which point the ball can exhibit resilience and return to a new local minima of health, or become associated with a local area minima of disease, and is unable to easily return to a healthy state. Using this framework, after antibiotic perturbation all 20 volunteer microbiomes are pushed out of the "healthy" state they were previously defined by and move through PCA space. Most volunteer microbiomes we tracked eventually return to a location within the original area of the PCA the healthy volunteer microbiomes started in. The three "ill-like" individuals do not, and instead inhabit the PCA space defined by the ICU microbiomes. I believe our data to be a confirmation of the validity of the model for microbiome dynamics mentioned by Lloyd-Price, though the small sample size of our study requires larger-scale follow up to fully characterize this phenomenon. Regardless, I believe that these data are an important step towards designing larger scale studies to expand the number of tested antibiotics, and alongside experimental modulation of diet, to test for interaction between microbiome recovery and diet in humans. With this work, I expand the scientific community's knowledge of the dynamics of gut microbiome remodeling and recovery in healthy volunteers and inform the design of future studies.

The work I conducted in chapter 2 of this dissertation is another experimental perturbation of the bacterial cell, though smaller in ecological scale. Altering *E. coli* for novel fatty acid composition is becoming a well-established field, yet the fundamental aspects of altering the phospholipid membrane (considered one of the most important components of the cell) are often overlooked in pursuit of producing a strain ready for commercialization. I utilized transcriptomic profiling of 4 bacterial strains to identify evidence of perturbation in *E. coli* during the generation of the novel DUFAs. Though the unsaturated FAs used in this study differ from saturated FAs by the addition of one or two double bonds, their overproduction has broad implications on the structure, fluidity, and packing of the membrane. Not surprisingly, phenotypic characterizations of strains overproducing UFAs and DUFAs revealed large decreases in metabolic activity in many environmental conditions, while cyclopropane FAs (CFA) and internally branched-chain fatty acids (IBFA) were largely able to reduce effects on metabolism.

Interestingly though, this pattern was not seen when comparing growth rates, or transcriptional profiles. This indicates that the detrimental effects of UFA production are not due to altered transcriptional regulation or energy demand, but rather by changes to membrane structure. The DUFA producing strain alone created large scale perturbations to the transcriptome, resulting in high differential expression of almost 400 genes. As further evidenced by over 20 differentially expressed genetic regulatory genes, multiple regulatory regimes were altered. Induction of the *mar* and *sox* stress-related regulons indicate a heightened general stress response. Increased expression of iron-sulfur cluster containing proteins and iron-starvation can activate these genetic programs(12, 13), and we observe increased expression of the *fur regulon*. Increased iron-acquisition is likely necessary for our method of DUFA production which

requires overexpression of a novel ferredoxin. When taken together, the negative phenotypic and transcriptional alterations seen in DUFA signify multiple areas for further engineering. Deletion of *rhyB*, a small RNA which regulates *fur* itself can potentially thread the needle, eliminating *fur* induced restraints on central metabolism while still increasing expression of iron acquisition and utilization systems. By characterizing the phenotypic and transcriptomic alterations induced during the creation of novel phospholipid compositions I add important context to our knowledge of off-target effects of synthetic design on the bacterial chassis.

We are living in a warming world. Global climate change is, and will continue to be, the greatest challenge of our lifetimes(14). Microbial bioproduction is a promising avenue for degrading current waste streams into value added chemicals, but is hampered by production inefficiency due to low natural utilization rates and substrate tolerance(15). Developing the *Rhodococcus opacus PD630* chassis for greater utilization and tolerance on aromatic compounds will have a real impact on future methods for biofuel generation(16). In chapters 4 and 5 of my dissertation I use two different methods of bioengineering – evolutionary adaption and rational design – to identify regulatory mechanisms in *R. opacus* important for aromatic tolerance and lipid synthesis. The direct mechanisms for each of these have already been identified in *R. opacus;* and there is consensus in the field that future increases in either carbon storage molecule production or aromatic tolerance will entail modifying multiple metabolic pathways and balancing the resources shared between them(17, 18). This will almost certainly result in off target effects to metabolism and necessitate deeper multi-omic characterization to identify solutions. Towards a deeper characterization of the secondary mechanisms of regulating fatty acid content in *R.* opacus, we identified 3 novel regulators of fatty acid content, none of which were annotated as being related to fatty acid biosynthesis. Importantly, the normal metabolic

requirement for limited resources was circumvented by these novel regulatory elements, which successfully increased fatty acid content in nitrogen replete environments. Compositional analysis of the transcriptomic effects of each regulator revealed that ATR 13 and ATR 20, seemingly independent of one another, produced a similar response when grown in either phenol. This involved increasing expression of phenol catabolism and protein synthesis genes, as well as upregulation of multiple cofactor biosynthesis pathways. The importance of balancing the requirement for cellular resources is becoming more and more apparent in biological engineering(17, 19), and we can reasonably conclude that ATR13 increases phenolic catabolism via augmenting demand for available cell resources.

Strain 13, the strain with the highest lipid production titers, overexpresses the negative regulator of the phenylacetic acid (paa) degradation pathway. Further analysis of upstream components of phenylalanine metabolism uncovered a complicated regulatory mechanism requiring the activity of *feaR,* the phenylethylamine (pea) promoter. Pea degradation is directly upstream of the paa pathway, indicating that production of paa, but not degradation, is important for increased fatty acid synthesis. Future experiments interrogating this method for increasing carbon storage in nutrient replete conditions are necessary and will require identifying the mechanistic link between paa degradation and phenolic catabolism. Given the structural similarities between components of the paa and β-ketoadipate pathway (the main pathway for integrating aromatic degradation products into central metabolism), I hypothesize that uncoupling paa degradation allows for cofactor reallocation to phenolic degradation and catabolism.

Any future biofuel production using *R. opacus* as a chassis will need to overcome the toxicity of lignin breakdown products, which tend to degrade into aromatic and phenolic

derivatives(20, 21). The *R. opacus* degradation and utilization pathways for aromatic compounds are known(22-24), and strains have been successfully evolved for increased tolerance to model aromatic compounds found in lignin breakdown products(22). Previous work attempted to uncover the regulatory mechanisms responsible for increased tolerance(22). Transcriptomic analysis of the evolved strain producing the highest lipid titers was unable to identify differential expression in any of the aromatic catabolism pathways, and thus deeper characterization of the transcriptional differences after adaptation were required. To address this, I conducted RNA expression profiling of 3 strains adaptively evolved for increased tolerance to mixtures of aromatic lignin breakdown products (LBPs).

By analyzing the expression of each mutant challenged at a high concentration of MLBPs, I was finally able to observe differential expression of the β-ketoadipate pathway. Importantly, PVHG6, the strain adapted to the most complex MLBP did not upregulate the pathway as PV1 and PVH5 did, but instead appeared to partially decrease expression. This same divergence in expression was identified in a putative energy/lipid metabolism operon which was differentially expressed in all three strains. Similar divergent expression patterns have been detected in closely *Burkholderia* species after adaption to different niche requirements(25); we can conclude that divergent expression identified here in *R. opacus* strains is directly related to the adaption strategies employed by the different strains. Future work is needed to investigate the function of this set of genes, and to clarify the role downregulating the β-ketoadipate pathway plays in PVHG6's increased tolerance.

I hypothesized that growing *R. opacus* on increasingly complex mixtures of LBP (MLBPs) would result in transcriptional profiles compositionally similar to *R. opacus* grown in LBP; furthermore, adapted strains would be even more similar. Using WT *R. opacus* grown in

LBP as a direct comparison, I was able to refute this hypothesis: There was no appreciable relationship between increasing MLBP complexity and compositional similarity to WT grown in LBP. *R. opacus* grown in LBP also exhibited markedly different expression levels of the aromatic degradation and funneling pathways, indicating that adapting to growth on aromatics as sole carbon sources is not representative of the metabolic requirements when grown on real world lignin. This is an important negative result, as we can re-allocate resources away from adapting *R. opacus* on model LBPs and towards transcriptomic characterization of mutants adapted on real world lignin(26).

Overall, I conclude that "antibiotic scarring" after treatment with antibiotics occurs in the healthy human microbiome (chapter 2). Though antibiotic scarring may be unfelt by the host, it permanently alters the composition and resistome of the gut microbial community, and some individuals are primed for increased effects. Refocusing to the cellular level, I find that altering the composition of the phospholipid membrane results in a large off-target transcriptional response due to increased resource requirements (chapter 3). Similarly, successful reallocation of cellular resources and transcriptional remodeling results in increased lipid production in *R. opacus* (chapter 4). As we work to increase lignin tolerance in *R. opacus* (chapter 5), we must refocus our efforts away from model lignin breakdown products, which produce adaptation not representative of growth on real world lignin.

References

1. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proceedings of the National Academy of Sciences. 2011;108(supplement_1):4554-61.

2. Anthony WE, Wang B, Sukhum KV, D'Souza AW, Hink T, Cass C, et al. Acute and persistent effects of commonly used antibiotics on the gut microbiome and resistome in healthy adults. Cell Reports. 2022;39(2).

3. Dethlefsen L, Huse S, Sogin ML, Relman DA. The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. PLOS Biology. 2008;6(11):e280.

4. Edwards R. Resistance to β-Lactam Antibiotics in Bacteroides Spp. Journal of Medical Microbiology. 1997;46(12):979-86.

5. Ng KM, Aranda-Díaz A, Tropini C, Frankel MR, Van Treuren W, O'Loughlin CT, et al. Recovery of the Gut Microbiota after Antibiotics Depends on Host Diet, Community Context, and Environmental Reservoirs. Cell Host Microbe. 2019;26(5):650-65.e4.

6. Drew RH, Gallis HA. Azithromycin—Spectrum of Activity, Pharmacokinetics, and Clinical Applications. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy. 1992;12(3):161-73.

7. Fish DN, Chow AT. The Clinical Pharmacokinetics of Levofloxacin. Clinical Pharmacokinetics. 1997;32(2):101-19.

8. Borin MT. A Review of the Pharmacokinetics of Cefpodoxime Proxetil. Drugs. 1991;42(3):13-21.

9. Jacoby GA, Strahilevitz J, Hooper DC. Plasmid-mediated quinolone resistance. Microbiology spectrum. 2014;2(5):10.1128/microbiolspec.PLAS-0006-2013.

10. Rashidi A, Ebadi M, Rehman TU, Elhusseini H, Nalluri H, Kaiser T, et al. Gut microbiota response to antibiotics is personalized and depends on baseline microbiota. Microbiome. 2021;9(1):211.

11. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome Med. 2016;8(1):51.

12. Seo SW, Kim D, Latif H, O'Brien EJ, Szubin R, Palsson BO. Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in Escherichia coli. Nat Commun. 2014;5:4910-.

13. Hantke K. Iron and metal regulation in bacteria. Current Opinion in Microbiology. 2001;4(2):172-7.

14. Prasad PVV, Thomas JMG, Narayanan S. Global Warming Effects. In: Thomas B, Murray BG, Murphy DJ, editors. Encyclopedia of Applied Plant Sciences (Second Edition). Oxford: Academic Press; 2017. p. 289-99.

15. Ling H, Teo W, Chen B, Leong SSJ, Chang MW. Microbial tolerance engineering toward biochemical production: from lignocellulose to products. Current Opinion in Biotechnology. 2014;29:99-106.

16. Anthony WE, Carr RR, DeLorenzo DM, Campbell TP, Shang Z, Foston M, et al. Development of Rhodococcus opacus as a chassis for lignin valorization and bioproduction of high-value compounds. Biotechnology for Biofuels. 2019;12(1):192.

17. Borchert AJ, Henson WR, Beckham GT. Challenges and opportunities in biological funneling of heterogeneous and toxic substrates beyond lignin. Current Opinion in Biotechnology. 2022;73:1-13.

18. Linger JG, Vardon DR, Guarnieri MT, Karp EM, Hunsinger GB, Franden MA, et al. Lignin valorization through integrated biological funneling and chemical catalysis. Proceedings of the National Academy of Sciences. 2014;111:12013-8.

19. Akhtar MK, Jones PR. Cofactor Engineering for Enhancing the Flux of Metabolic Pathways. Frontiers in Bioengineering and Biotechnology. 2014;2.

20. Mottiar Y, Vanholme R, Boerjan W, Ralph J, Mansfield SD. Designer lignins: harnessing the plasticity of lignification. Current Opinion in Biotechnology. 2016;37:190-200.

21. Holladay JE, White JF, Bozell JJ, Johnson D. Top value-added chemicals from biomass-Volume II—Results of screening for potential candidates from biorefinery lignin. Pacific Northwest National Lab.(PNNL), Richland, WA (United States); 2007.

22. Henson WR, Campbell T, DeLorenzo DM, Gao Y, Berla B, Kim SJ, et al. Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. Metabolic engineering. 2018;49:69-83.

23. Yoneda A, Henson WR, Goldner NK, Park KJ, Forsberg KJ, Kim SJ, et al. Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. Nucleic Acids Res. 2016;44(5):2240-54.

24. Hollinshead WD, Henson WR, Abernathy M, Moon TS, Tang YJ. Rapid metabolic analysis of Rhodococcus opacus PD630 via parallel 13C-metabolite fingerprinting. Biotechnology and Bioengineering. 2016;113(1):91-100.

25. Kim HS, Schell MA, Yu Y, Ulrich RL, Sarria SH, Nierman WC, et al. Bacterial genome adaptation to niches: Divergence of the potential virulence genes in three Burkholderia species of different survival strategies. BMC Genomics. 2005;6(1):174.

26. Kurosawa K, Laser J, Sinskey AJ. Tolerance and adaptive evolution of triacylglycerol-producing Rhodococcus opacus to lignocellulose-derived inhibitors. Biotechnology for Biofuels. 2015;8(1):76.

Curriculum Vitae

WINSTON EUGENE ANTHONY

Mobile phone • 978-394-5100 • winstoneanthony@wustl.edu

Summary/Skills

Computational microbiology Ph.D. candidate with 6 years of research experience in biological data science. Expertise in the design of HPC pipelines for shotgun whole genome metagenomic and comparative transcriptomic analysis. Proficient in an array of bioinformatics software packages, operating systems, and languages. Experience in leading a team of scientists to complete projects funded by the CDC and DOE.

- **Molecular Biology:** Metagenomics (DNAseq), Transcriptomics (RNAseq), Bacterial cell culture, *C. elegans* culture
- **Languages:** R (advanced proficiency), Bash (as needed), Python (as needed)
- **Packages: R** (ggplot2, e1071, vegan, dplyr, stringr, reshape2, tidyr), **Python** (Bio.Blast, bs4)
- **Operating Systems:** Ubuntu Linux, SLURM Linux Mac OS, Windows
- **Bioinformatic Skills:** Bowtie2, SAMtools, FeatureCounts, TRIMMOMATIC, Metaphlan, HMMER
- **Cloud Computing:** Amazon web services (AWS), Docker, git version control, R Shiny
- **Interpersonal:** Excellent communicator, Strong leadership, Dependable in teamwork setting, Creative when presented with difficult problems.

Education

Ph.D. in Molecular Cell Biology, Washington University in St. Louis, MO, August 2022 (expected)
Dissertation Title: Emergent Properties of the Taxa-Function Landscape of Bacteria
B.S. in Biology, University of Massachusetts at Amherst**,** MA, 2013
Certificate in Marine Science, University of Hawaii at Hilo, HI, 2011

Research Experience

Antibiotic perturbation of human gut microbiomes, Advisor: Gautam Dantas Feb 2017-Present
Washington University in St. Louis

- CDC funded study of clinical cohort given antibiotics and followed longitudinally for 6 months.
- Analyzed semi-quantitative culture data to detect loss of live bacteria in human microbiomes after antibiotics.
- Extracted gDNA and sequenced 289 fecal metagenomes using whole-genome shotgun DNAseq, along with 200+ isolate genomes.
- Conducted metagenomic analysis to discover biomarkers of microbiome dysbiosis after antibiotics in healthy volunteers. Analysis of bacterial functional output and antibiotic resistance load after antibiotics.
- Research articles published in *Cell Reports* and press coverage in The Scientist.

Differential expression for altered lipid production and aromatic tolerance, Advisor: Gautam Dantas

- DOE funded study generating bacterial mutants for biofuel production and lignin valorization
- Extracted mRNA and sequenced 110 bacterial transcriptomes. In-house development of custom linux HPC pipeline for RNAseq expression profiling.
- Supervised a team of researchers to coordinate experimental design, processing, and analysis of projects.
- Transcriptomic analysis to uncover differentially expressed genes during triacylglyceride production in *Rhodococcus opacus*
- RNAseq, growth rate, and biomass analysis of mutant *Echerichia coli* designed for novel lipid compositions.

Data science internship, Summer 2021
Pivot Biosciences

- Developed an Amazon Web Services (AWS) pipeline for detecting differential gene abundance in soil metagenomes
- Learned secure data management practices by implementing a multi-node parallel batch job structure and integrated multiple tools into a multi-container application stack.

Temporal scaling of lifespan in *C. Elegans*, Principal Investigator: Walter Fontana**,** 2013-2016
Harvard Medical School

- Independently performed large scale experiments investigating *C. elegans* thermotolerance and lifespan using an automated image-analysis device developed in-house.
- Research articles published in *Nature* and press coverage in STAT and Science Daily.

Bioactive chemical screens in pathogen infected *B. impatiens*, Advisor: Lynn Adler, 2011-2013
University of Massachusetts at Amherst

- Conducted a drug screen analyzing the effect of nectar alkaloids on Crithidia bombi levels in common eastern bumblebee *B. impatiens*.
- Research Articles published in *Proc. Royal Society B* and *PLoS One*.

DNA Barcoding to detect Seafood substitution, Advisor: Luiz Rocha
University of Texas at Austin

- DNA isolation/purification of tissue samples from consumer fish fillets. Refined PCR protocols to target specific fish species. Utilization of DNA extraction, PCR, and gel electrophoresis.

Research Articles

- Bai W[†], **Anthony W.E**[†], Hartline C, Shaojie W, Wang B, Ning J, Hsu FF, Dantas G, Zhang F. Engineering diverse fatty acid compositions of phospholipids in Escherichia coli. *Metabolic Engineering,* (2022).
- **Anthony W.E.**, Wang B, Cass C, Hink T, Reske K, Seiler S, Dubberke ER, Burnham CA, Dantas G, Kwon JH. Acute and persistent effects of commonly-used antibiotics on the gut microbiome and resistome in healthy adults. *Cell Reports* 39.2 (2022): 110649.
- **Anthony, W.E.**, Burnham, C.A.D., Dantas, G., Kwon, J.H., 2020. The Gut Microbiome as a Reservoir for Antimicrobial Resistance, *The Journal of Infectious Diseases*, 2020; jiaa497
- **Anthony, W.E.**, Carr, R.R., DeLorenzo, D.M., Campbell, T.P., Shang, Z., Foston, M., Moon, T.S. and Dantas, G., 2019. Development of Rhodococcus opacus as a chassis for lignin valorization and bioproduction of high-value compounds. *Biotechnology for biofuels*, *12*(1), pp.1-14.
- Stroustrup, N., **Anthony, W.E.**, Nash, Z.M., Gowda, V., Gomez, A., López-Moyado, I.F., Apfeld, J. and Fontana, W., 2016. The temporal scaling of Caenorhabditis elegans ageing. *Nature*, *530*(7588), p.103.

- **Anthony, W.E.**, Palmer-Young, E.C., Leonard, A.S., Irwin, R.E. and Adler, L.S., 2015. Testing dose-dependent effects of the nectar alkaloid anabasine on trypanosome parasite loads in adult bumble bees. *PloS one*, *10*(11), p.e0142496.
- Richardson, L.L., Adler, L.S., Leonard, A.S., Andicoechea, J., Regan, K.H., **Anthony, W.E.**, Manson, J.S. and Irwin, R.E., 2015. Secondary metabolites in floral nectar reduce parasite infections in bumblebees. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1803), p.20142471.
  †Equal Contribution

Press, Fellowships, and Awards

- Fessl, S. (2022, May 5). What happens to the gut microbiome after taking antibiotics. The Scientist Magazine®. Retrieved May 13, 2022
- Pivot 314 Fellow (2021)
- Top Oral Presentation (2020) – 6th Decennial International Conference on Healthcare Associated Infections
- Outstanding Abstract Award (2019) – American Society of Microbiology annual meeting, San Francisco, CA
- T32 NIH Ruth L. Kirschstein National Research Training Grant Fellowship – Cellular and Molecular Biology: 5T32GM007067-44 (2017-2019)
- Initiative for Maximizing Student Development (IMSD) (2016 – 2018)
- Harvard Medical School Science Communications. The Lifespan Machine. Youtube**.**
- Harvard Medical School Research Fellowship – Laboratory of Walter Fontana, Ph.D (2013 – 2015)
- National Science Foundation Summer Research Experience for Undergraduates – UT Austin REU in Subtropical Marine Ecosystems (2012)
- Concord Carlisle Scholar Fund Awardee (2011)

Presentations

Anthony, W.E., Diao J., Roell G., Hu Y., Carr R., Davis K., DeLorenzo D., Wang B., Ning J., Foston M., Zhang Fuzhong, Martin H.G., Tang Y., Moon T.S., Dantas G. Elucidating aromatic utilization mechanisms in engineered *Rhodococcus opacus* strains for lignin valorization. Department of Energy Genomic Sciences Program Annual Meeting, Washington, DC. February 2021

Anthony, W.E., Wang, B., Cass, C., Hink, T., Reske, K., Seiler, S., Dubberke, E.R., Reske K.A., Burnham, C.D., Dantas, G., Kwon, J.H. Perturbation of the healthy human microbiome after antibiotics is remediated through functional resiliency and increased resistance. American Society of Microbiology annual meeting, San Francisco, CA. June 2019.

Anthony, W.E., American Society of Microbiology annual meeting Poster Talk, San Francisco, California, "Machine Learning Identifies a Subset of the Microbiota Which Defines Perturbation to the Fecal Microbiome of Healthy Humans" June 2019.

Anthony, W.E., Wang, B., Cass, C., Hink, T., Reske, K., Seiler, S., Dubberke, E.R., Reske K.A., Burnham, C.D., Dantas, G., Kwon, J.H. Machine Learning Identifies a Subset of the Microbiota Which Defines Perturbation to the Fecal Microbiome of Healthy Humans. Lake Arrowhead Microbial Genomics Meeting, Lake Arrowhead, CA. September 2018.

Anthony, W.E., Carr, R., Campbell, T., Henson, W.R., Moon, T.S., Dantas, G. Characterizing the transcriptomes of Rhodococcus opacus strains adapted to model lignin breakdown products. Department of Energy Genomic Sciences Program Annual Meeting, Washington, DC. February 2018

Leadership and Student Organizations

Student Leadership Committee
Washington University in St. Louis

- Roles: Secretary, Division of Biological and Biomedical Sciences Graduate Student Senate Representative

WashU Medical School Musical
Washington University in St. Louis

- Ensemble member: 2017-2018, 2018-2019; Assistant Producer (Chicago!): 2019-2020

Mentoring Experience

Mentored Teaching Experience (MTE) for WUSTL BIO 2960, Principles of Biology I
Washington University in St. Louis

- Taught laboratory component for DNA sequence data analysis for >30 WUSTL undergraduate students

Nashwa M Ahmed, B.S. – Washington University in St. Louis undergraduate student Fall 2019 – Spring 2020. Mentored for a senior thesis project: Analyzing single nucleotide polymorphism distance between strains of *E. coli*. Advised during Ph.D. interview season which was ultimately successful.
Tenia Sims, B.S. – YSP Summer Focus high school student at WUSTL Summer 2017. Tutored in biological concepts, scientific thinking/theory, and in applying for colleges