THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

GENOME-WIDE SCREEN FOR THE INDUCTION OF PHOSPHOETHANOLAMINE AND AMINOARABINOSE MODIFICATIONS ON *ESCHERICHIA COLI* LIPOPOLYSACCHARIDE

ALEXANDER SMITH

SPRING 2018

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biochemistry and Molecular Biology
with honors in Biochemistry and Molecular Biology

Reviewed and approved* by the following:

Timothy Meredith
Assistant Professor of Biochemistry and Molecular Biology
Thesis Supervisor

Wendy Hanna-Rose
Associate Professor and Interim Department Head of Biochemistry and Molecular Biology
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Chronic inflammation arising from metabolic endotoxemia (ME) is currently a widespread health concern with deep ties to the complex human microbiome. Gram-negative microbiota such as *Escherichia coli* possess lipopolysaccharide (LPS), a potent endotoxin and mediator of this chronic inflammation. Covalent modification of LPS with phosphoethanolamine (PEtN) has been shown to exacerbate this response, so understanding the genetic regulation of PEtN modifications is essential to fully understanding ME. Different strains of *E. coli*, however, have a spectrum of basal PEtN expression despite there being 100% DNA sequence conservation within the PEtN modifying gene promoter itself. Methods to identify these strain specific alleles are thus needed. We have constructed a two-component genetic reporter system in *E. coli* K-12, a strain normally silent for LPS modification, that allows us to introduce and positively select transposon linked alleles from other strains of *E. coli* that specifically activate this modification pathway. By coupling PEtN activation to cell growth, we have used next-generation sequencing analysis to identify several putative alleles in strain *E. coli* BL21(DE3) that are responsible for constitutive expression of PEtN modifications. The reporter system that we have developed here can not only be applied to rapidly analyze any *E. coli* microbiome isolate in order to uncover specific PEtN regulatory alleles, but also to other promoter elements of interest.

TABLE OF CONTENTS

LIST OF FIGURES iii

LIST OF TABLES iv

ACKNOWLEDGEMENTS v

Chapter 1 Introduction 1

Chapter 2 Reporter System Design and Construction 6

- Background 6
- Results 9
- Methods 14

Chapter 3 Reporter System Validation 17

- Background 17
- Results 18
- Discussion 21
- Methods 22

Chapter 4 Transposon Library Construction and Transduction 25

- Background 25
- Results 26
- Discussion 33
- Methods 34

Chapter 5 Next-Generation Sequencing 36

- Background 36
- Results 37
- Discussion 39
- Methods 41

Chapter 6 Conclusion and Future Directions 46

Appendix Supplementary Information 50

BIBLIOGRAPHY 53

LIST OF FIGURES

Figure 1: TLR pathway. 2

Figure 2: LPS structure 3

Figure 3: NF-κB induction 4

Figure 4: Reporter system schematic. 7

Figure 5: Full experimental workflow. 8

Figure 6: pSEVA plasmid map 10

Figure 7: Tn7 delivery system. 12

Figure 8: β-galactosidase activity assay. 19

Figure 9: Curing assay example 20

Figure 10: Curing ratio validation 21

Figure 11: Plasmid maps of transposon library generation system 27

Figure 12: Kanamycin-resistant cell count during outgrowth 29

Figure 13: Chloramphenicol-resistant cell count during outgrowth 30

Figure 14: Transduction control cell count during outgrowth 31

Figure 15: Next-generation sequencing alignments 37

Figure 16: Significant transposon insertion sites 38

Figure 17: Hypotheses for phenotypic divergence 47

LIST OF TABLES

Table 1: *TnpR* Allele Plating Assay 12

Table 2: Final transduction cell count data in MYAS510 32

Table 3: Putative causative alleles of differential phenotype 41

Table 4: List of Strains 50

Table 5: List of Plasmids 51

Table 6: List of Primers 52

ACKNOWLEDGEMENTS

I would like to thank Dr. Tim Meredith for his guidance and mentorship on matters both inside and outside the lab over the past four years. He has been an integral component of my undergraduate education. I would also like to thank Michael Yi, my partner on this independent research project, for his dedication and contribution. A special thanks to Gloria, Krista, Kelvin, Michael, and all members of the Meredith Lab for being an incredible support system and my home away from home at Penn State. Finally, I would like to thank my amazing parents, Jan and Tom Smith, and all of my close friends at Penn State for all of the love, support, and advice that got me through my four years here and prepared me for life beyond.

Chapter 1

Introduction

The complex human microbiome is currently a subject of intense scientific scrutiny because research in this area could lead to actionable, personalized treatment solutions for patients with as-of-yet untreatable illnesses. The constituents of this community as well as its diversity have been discovered to play a role in human health. Characteristics of the host-microbiome relationship have been implicated in chronic issues like diabetes, obesity, and resultant cardiovascular diseases, the latter of which is predicted to claim the lives of over 23 million people by 2030 (1, 2). Gram-negative proteobacteria such as *Escherichia coli*, common inhabitants of the gut microbiome, possess an outer-membrane leaflet composed of lipopolysaccharide (LPS), also called endotoxin (3). LPS is one of several bacterial products, such as peptidoglycan, lipoteichoic acid, and flagellin, originating from the gut microbiota that can act as pathogen-associated molecular patterns (PAMPs) to induce an innate immune response (4, 5). The intestinal epithelia normally exclude these PAMPs from the bloodstream. Yet, evidence suggests that a high-fat diet can induce increased permeability of the intestinal wall, leading in turn to increased leaching of LPS into the bloodstream (4, 6). There, the LPS can interact with the Toll-like receptor 4 (TLR4) complex, a receptor in the innate immune system, specifically evolved to recognize the highly conserved structure of the Lipid A motif of LPS. After binding of TLR4 to LPS and receptor oligomerization, TLR4 stimulates the production of downstream signal pathways that ready the body for infection. One of the major modes of activation is through NF-κB, a transcription factor

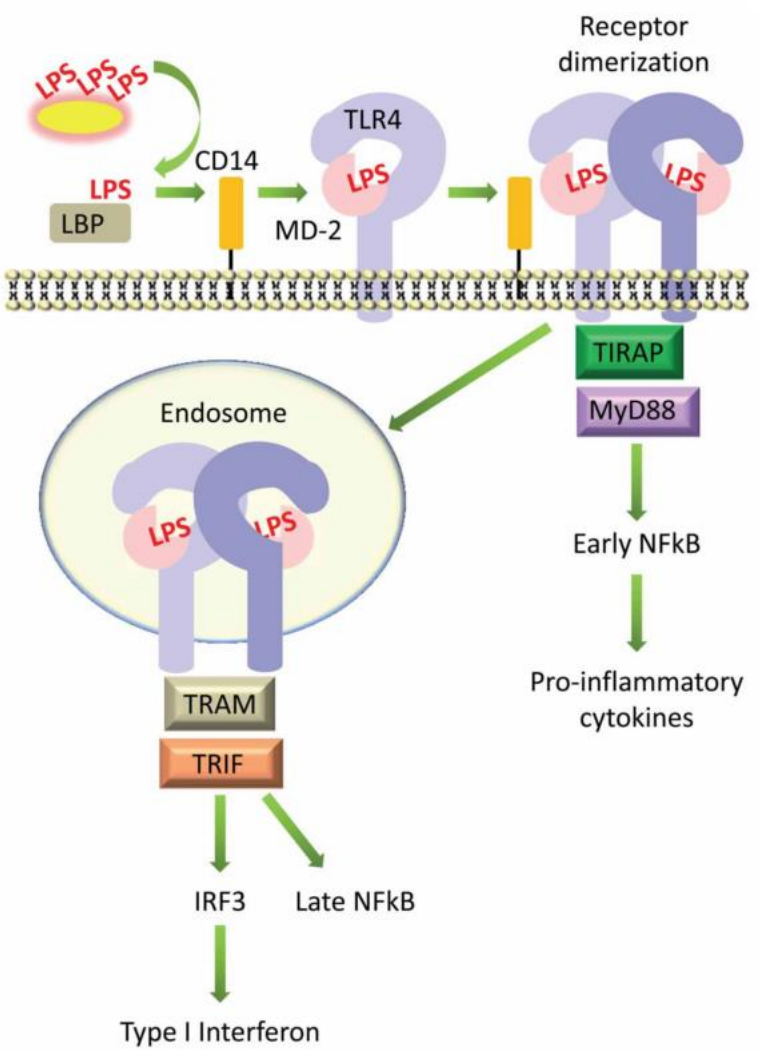


Figure 1: TLR pathway.

An overview of the TLR4 signaling pathway that is able to activate downstream pro-inflammatory and immunological signals through several mediators by binding LPS

that goes on to activate proinflammatory cytokines and Type 1 interferons (7, 8) (Figure 1). In the case of increased intestinal permeability and increased LPS leaching, the TLR4 sensors can be overstimulated to induce a chronic, low level of inflammation (9, 10). This signal interaction results in a chronic, low-level of inflammation that can cause increased oxidative stress and organ damage, including, but not limited to, cardiovascular disease, as mentioned above. This chronic

inflammation caused by high serum LPS levels is a condition called metabolic endotoxemia (ME) (1, 10).

E. coli is capable of differentially and covalently modifying its LPS, specifically the immune-reactive Lipid A component, with positively-charged functional groups. These modifications include phosphoethanolamine (PEtN) and aminoarabinose (L-AraN), added by the inner membrane proteins, *eptA* and *arnT*, respectively (Figure 2). These genes are under the

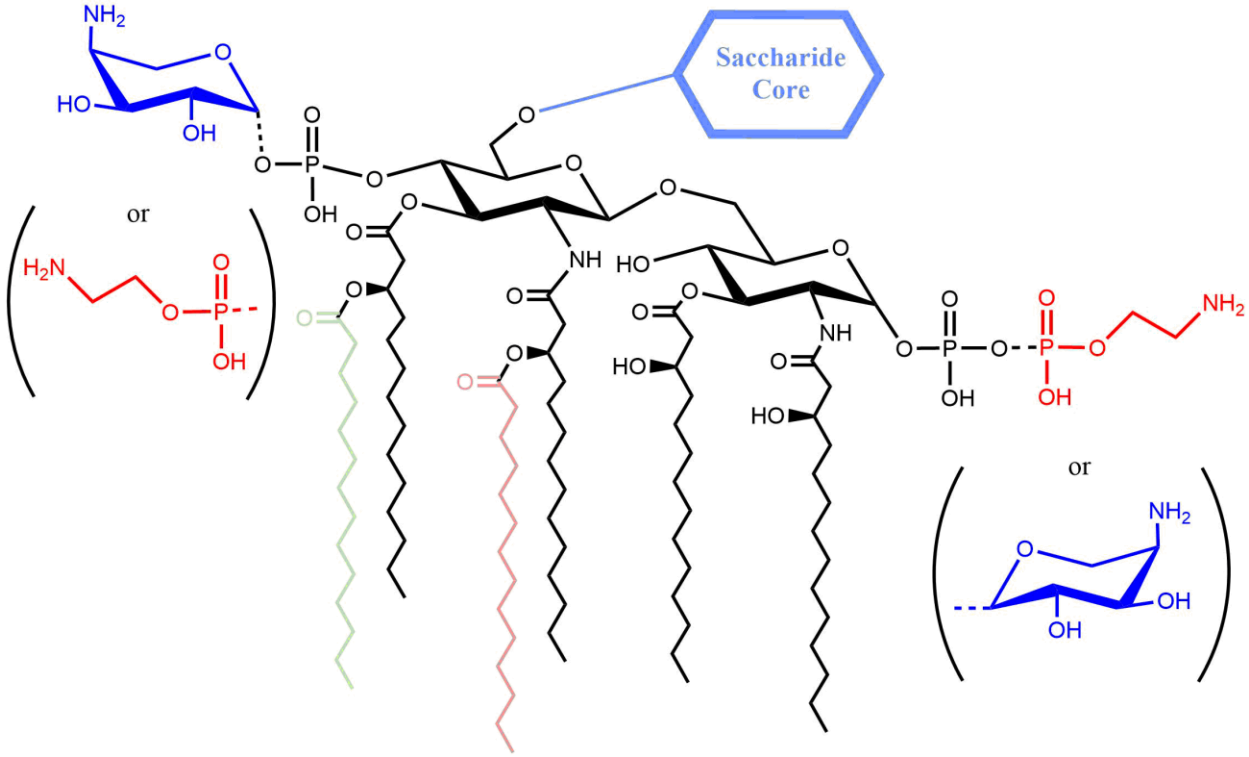
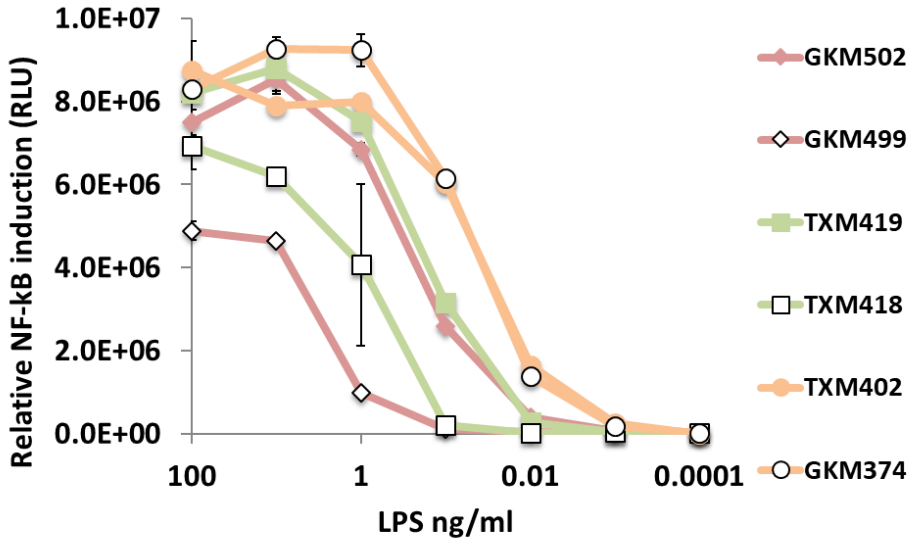


Figure 2: LPS structure

The Lipid A portion of LPS showing the possible sites of modification addition. PEtN (red) and L-AraN (blue) can be added onto either phosphate in any combination

control of the *basSR*, also called *pmrAB*, two-component system which activates these modifications when the bacterium is exposed to iron (Fe^{3+}), mildly acidic pH, aluminum, vanadate, and growth-limiting magnesium (11, 12). The sensing protein, *basS*, detects these periplasmic metal ions, Fe^{3+}most strongly, with an ExxE amino acid motif on a periplasmic domain. *BasS*

autophosphorylates and then phosphorylates the transcription factor *basR* to engage LPS-modifying genes (12). The system also responds to cationic antimicrobial peptides whose efficacy are limited by neutralizing the negative surface charge of the bacterium with these positively



Strain	PEtN	Acylation State
GKM499	no	tetra
GKM502	yes	tetra
TXM418	no	penta
TXM419	yes	penta
GKM374	no	hexa
TXM402	yes	hexa

Figure 3: **NF-κB induction**

The data above demonstrate the relative NF-κB induction over increasing LPS concentration for various constructs of LPS. The closed shapes represent PEtN modified LPS and the open shapes represent unmodified LPS. The colors correspond to acylation states of the LPS. As LPS is digested, deacylases remove acyl chains before it is excreted. The disparity of induction between modified and unmodified LPS is most notable at lower acylation states, suggesting that digestion could actually increase the problem.

charged moieties (13). It was shown that the PEtN and L-AraN modifications confer resistance to polymyxin B, currently a clinically-important antibiotic of last resort (11, 14).

My lab group has determined that the PEtN modified LPS ligands dramatically increase TLR4 receptor signaling and inflammatory cytokine secretion via NF-κB, particularly in under-acylated LPS chemotype variants (Figure 3). We hypothesize that exacerbation of this inflammatory response through a highly PEtN-modified LPS population can further worsen the symptoms of ME. Additionally, my lab has determined through mass spectrometric analyses of LPS isolates from different *E. coli* strains that endogenous PEtN modification levels are highly variable. *E. coli* BL21(DE3) (strain B lineage) was found to have an abundance of the above modifications while *E. coli* BW25113 (K-12) was found to have virtually none (15). Comparative analysis of the promoter regulatory regions of the *eptA* and *arnT* genes between K-12 and B revealed no significant genetic differences. Currently, no other global genetic regulators of this system are known. Understanding the regulation of, and therefore the circumstances of induction for, these LPS modifications is critical to understanding the pathology of ME. Elucidation of this regulatory mechanism could pave the way toward understanding how microbiome diversity contributes to human health. With this information, new treatments options may be available for patients with ME.

Chapter 2

Reporter System Design and Construction

Background

In order to elucidate the genetic elements responsible for this regulation, I have developed an *in vivo* recombination reporter system in an *E. coli* K-12 background that will detect induction of LPS modifications**.** My reporter system is a fusion of two previously developed reporters: recombinase-based in vivo expression technology (RIVET) and selectable in vivo expression technology (SIVET) (16, 17). The RIVET portion of the system consists of the promoter for our genes of interest coupled to the phage resolvase gene whose product removes DNA between flanking resolution sites. Once the *eptA* or *arnT* promoter is activated by an external growth condition, mutation, or acquisition of an inducing allele through transduction of foreign DNA (i.e. from *E. coli* B), the strain expresses resolvase which can then interact with SIVET. SIVET consists of a nonfunctional chloramphenicol resistance gene that has been interrupted by a tetracycline resistance gene flanked by resolution sites. These sites interact with the resolvase protein to remove tetracycline resistance and restore chloramphenicol resistance *only if* the *eptA*/*arnT* driven RIVET promoter is activated (Figure 4). To this reporter, I will

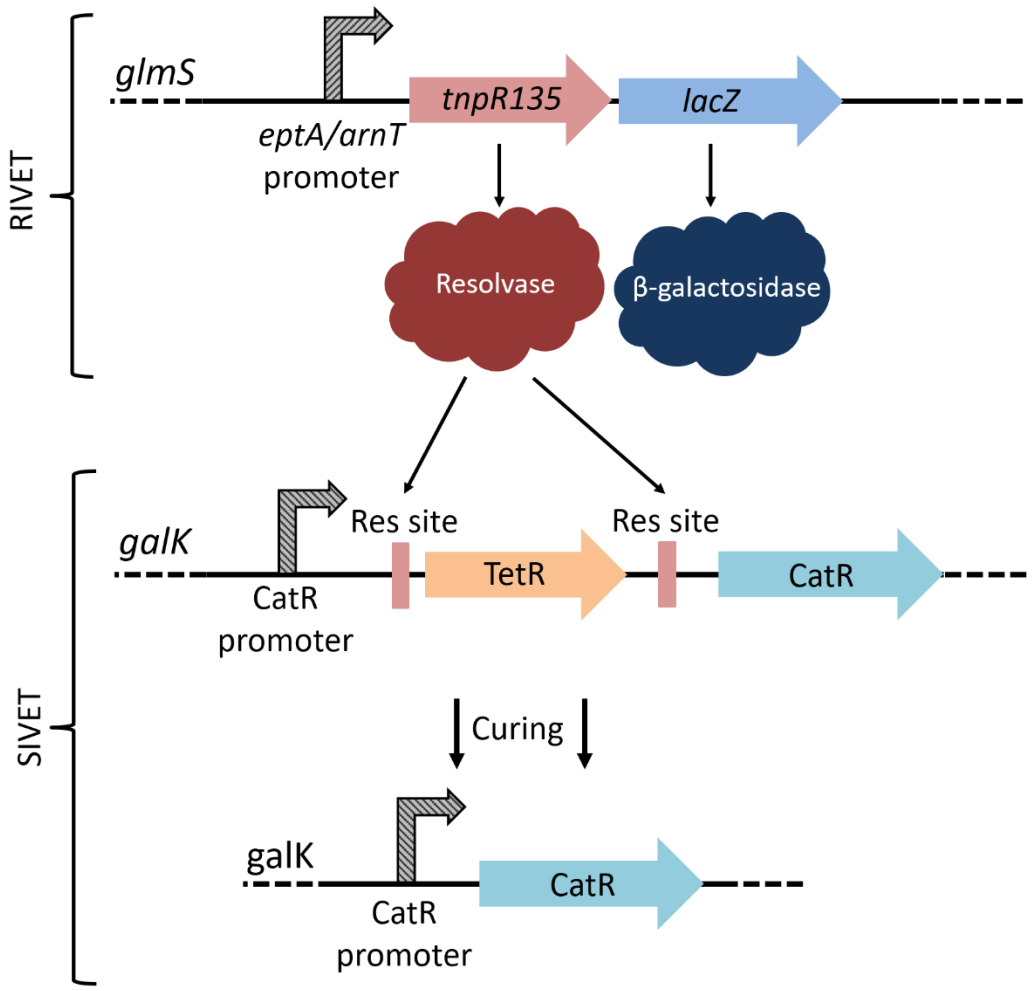


Figure 4: Reporter system schematic.

A workflow diagram demonstrating the interaction between the RIVET cassette, located at the *glmS* locus, and the SIVET cassette, located in the *galK* locus. Construction of this system is covered in Chapter 2 while validation is covered in Chapter 3.

transduce random segments of DNA from a high-density transposon library constructed in *E. coli* B. We hypothesize that there exist genetic regulatory elements in *E. coli* B that are capable of activating the promoters for modification when inserted into strain K-12. Colonies that have both a transposon from strain B and an activated reporter will be sequenced by next-generation

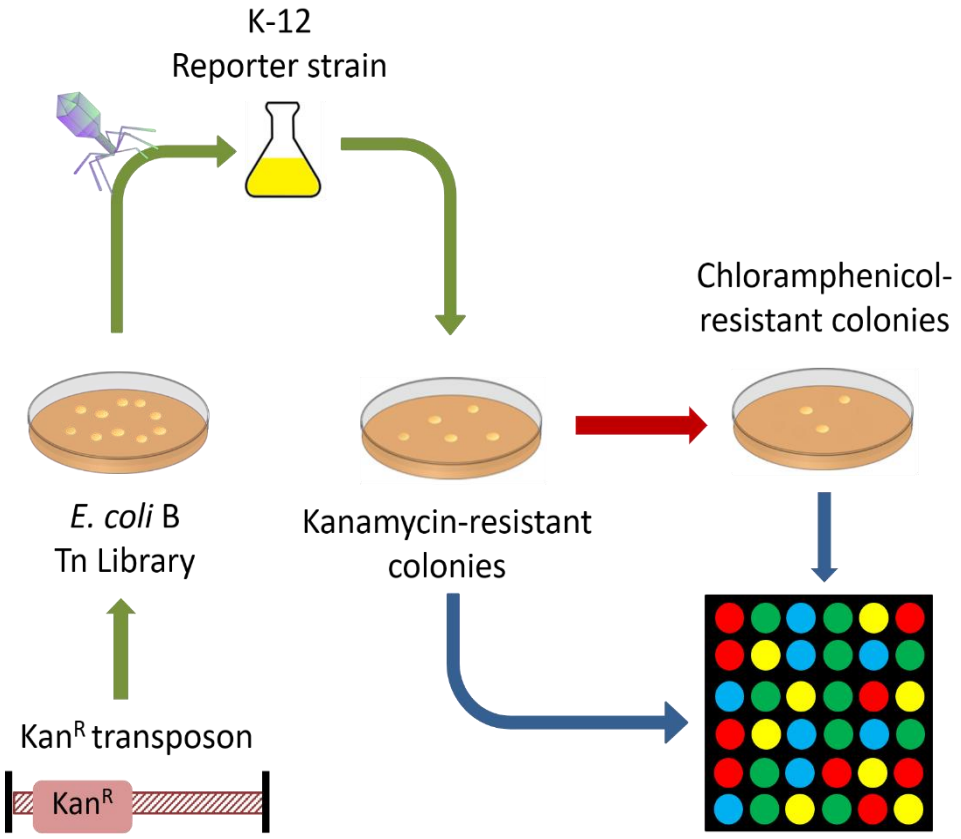


Figure 5: Full experimental workflow.

A full workflow diagram of the experimental procedure followed herein. First, a transposon library was generated in *E. coli* B. This library was then transduced via P1vir into our *E. coli* K-12 reporter strain (Chapter 4). After selection for the transposon, the reporter library was then plated on chloramphenicol to select for those bacteria that have an activated reporter. Genomic DNA was obtained from the unchallenged reporter library and the chloramphenicol-challenged library and submitted for next-generation sequencing (Chapter 5).

sequencing (NGS) in order to determine the global genetic elements in *E. coli* B that are responsible for the constitutive expression of LPS modifications (Figure 5). Due to the digital nature of the reporter (i.e. chloramphenicol resistance/sensitive), we propose that this system will be able to identify weakly activating alleles. This offers a tangible benefit over classical gene

expression reporters, such as green fluorescent protein (GFP) and β-galactosidase (*lacZ*), since the detection limit is not subject to assay sensitivity. A chloramphenicol-resistant colony is detected with the same effectiveness whether it arose from a strongly or weakly activating allele. Rather, the frequency of chloramphenicol resistance arising within the population dictates detection levels.

Results

First, we received strain K10447 from Dr. David Friedman (University of Michigan) which we then stocked as MYAS365. This strain, part of Friedman's SIVET system, contains the chloramphenicol resistance (*catR*) gene bisected by the tetracycline resistance (*tetR*) gene with flanking resolution sites, *resC*. This cassette sits directly in front of the *catR* gene promoter (7). The entire cassette is located in the *galK* locus. The SIVET cassette was amplified with primers AS553 and AS554 via PCR with NEB's Q5® High-fidelity DNA polymerase (Q5 polymerase PCR). The purified SIVET fragment was then electroporated into *E. coli* BW25113 (K-12) containing the temperature-sensitive pKD46 plasmid which allows homologous recombination of DNA fragments by phage λ Red recombinase (8). This new strain, *E. coli* K-12 with the SIVET cassette, was dubbed MYAS382.

Next, we received plasmids from Dr. Andrew Camilli (Tufts University) that contained portions of his RIVET 2.0 system. Plasmids pGOA1193, pGOA1194, and pGOA1195 contained the oriR6K *pir* product-dependent origin of replication, ampicillin resistance via *bla*, and the transcriptionally-coupled, promoterless resolvase (*tnpR*) and β-galactosidase (*lacZ*) genes. The plasmids contained different versions of *tnpR* with mutations in the ribosomal binding site

(RBS)—wild-type (TTTAGGA), mut168 (TTTAAGA), and mut135 (TTTGAGA), respectively—that allow the tuning of their expression level (6). Primers AS561 and AS562 were used to amplify the *tnpR-lacZ* cassette directly from the plasmids via Q5 polymerase PCR. Likewise, the *eptA* and *arnT* promoters were amplified from *E. coli* K-12 wild-type genomic DNA with primers TM466/467 and TM450/465, respectively. I digested a pUC19 plasmid vector, the promoter segments, and the *tnpR-lacZ* cassettes with compatible restriction enzymes for end joining and then coupled the three pieces with a three-piece, intermolecular ligation. This resulted in six unique plasmids with each of the two promoters joined to each of the three versions of *tnpR* (dubbed RIVET cassette). These plasmids were named pASCwt, pASC168, pASC135, pASHwt, pASH168, and pASH135. These plasmids were then electroporated into Stellar™ competent *E. coli* cells and dubbed MYAS389, 390, 391, 392, 393, and 394, respectively.

We performed a series of plating experiments in order to determine which of the *tnpR* alleles was the most effective. If the resolvase gene is too responsive, then there will be background activation of our reporter. Conversely, if the resolvase gene is not responsive enough, then we

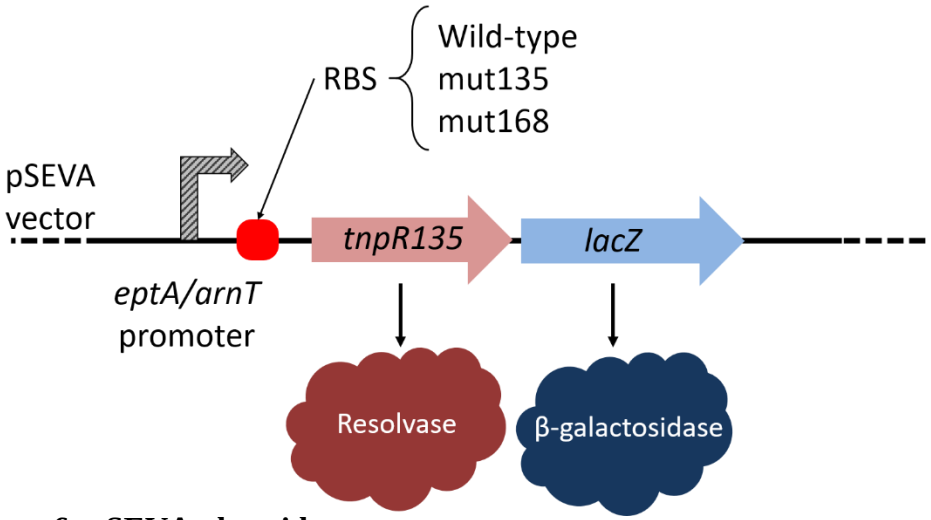


Figure 6: pSEVA plasmid map

The significant portion (RIVET) of the pSEVA plasmids transformed into the SIVET-containing strain of *E. coli* K-12 to generate MYAS436, 437, and 439 with emphasis on the various RBSs.

could potentially lose activating alleles in the final experiment. Each of the three *eptA-tnpR* (wild-type, mut135, and mut168) cassettes were cloned into a pSEVA vector and transformed into MYAS382—the SIVET-containing strain of *E. coli* K-12—in order to generate strains MYAS436, 437, and 439, respectively (Figure 6). Each of these three strains were grown in either inducing or non-inducing media designed to endogenously activate the *eptA* promoter. These formulations were developed according to the environmental factors that are known to regulate the *basSR* system (11, 12). For inducing conditions, we employed an N-minimal media with an acidic pH, limiting Mg^{2+} concentration, and a high Fe^{2+} concentration. For non-inducing, we employed an N-minimal media with a basic pH and a limiting concentration of Mg^{2+} (18). The strains were then plated on either LB agar, LB agar with tetracycline (5 µg/mL) (Tet5) or LB agar with chloramphenicol (5 µg/mL) (Cat5) to select for inactivated or activated reporters. The experiment was repeated in triplicate, the cell counts were normalized, and the data was averaged (Table 1). The curing ratio, or the number of chloramphenicol-resistant colonies over the sum of tetracycline and chloramphenicol-resistant colonies was calculated (Table 1). We ruled out mut168 because the RBS was non-functional and the gene was never able to express resolvase even under inducing conditions. The wild-type version possessed a higher background of expression under non-inducing conditions than mut135 which possessed a desirable balance between high curing under inducing conditions and low curing under non-inducing conditions. We ultimately chose to proceed with *tnpR*135 in our reporter constructs.

Table 1: *TnpR* Allele Plating Assay

	Inducing Media (CFU)		
	MYAS436 (wt)	**MYAS437 (135)**	**MYAS439 (168)**
Total (LB)	9.21E+09	6.72E+09	7.92E+09
Tet5	6.67E+06	2.43E+07	6.96E+09
Cat5	8.18E+09	6.49E+09	0.00
Curing Ratio	0.999	0.996	0.000

	Non-inducing Media (CFU)		
	MYAS436 (wt)	**MYAS437 (135)**	**MYAS439 (168)**
Total (LB)	5.76E+09	3.86E+09	3.60E+09
Tet5	8.73E+08	3.42E+09	6.61E+09
Cat5	2.66E+09	1.19E+08	0.00
Curing Ratio	0.753	0.034	0.000

We tried a variety of methods and neutral loci insertion sites for the RIVET cassette in MYAS 382, the *galK*::SIVET-containing strain. Since, unlike the SIVET cassette, the RIVET

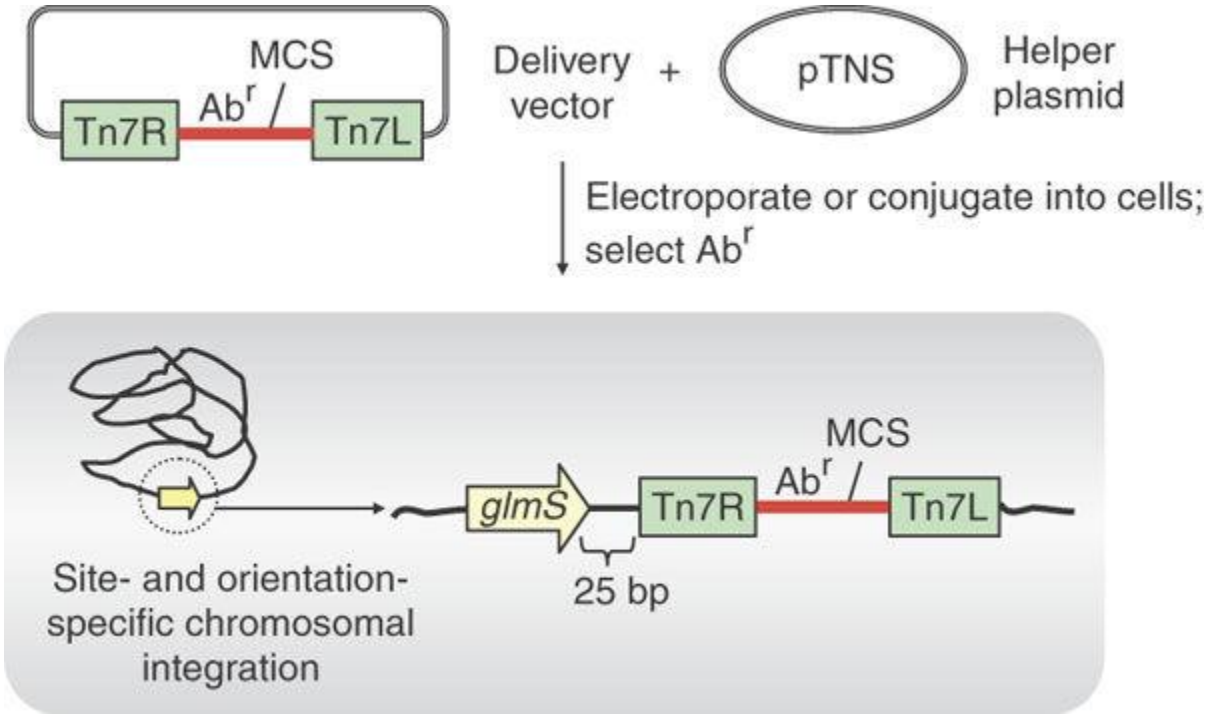


Figure 7: Tn7 delivery system.

An overview of the delivery system designed by Choi et al. that we used to deliver the promoter-resolvase cassette into the SIVET-containing strain.

cassette contains no selectable markers, we attempted to replace a pre-installed gentamicin-resistance gene in order to impart negative selection for transformation. Even at different loci, this method failed. We ultimately employed a Tn7-based transposition method developed by Choi et al. (19, 20) (Figure 6). This method relies on the delivery of two *pir*-dependent suicide plasmids, one of which contains the cassette to be inserted and the other, the transposase. When both are transformed into the target strain, the site-specific transposase will transfer the cassette into the *glmS* locus. The plasmids are then unable to replicate because they are no longer in a pir^+ strain. I received the two plasmids, pTNS2 (the transposase plasmid) and pUC18R6K-mini-Tn7T-Gm (the delivery plasmid), and electroporated them into the Transformax™ EC100D™ *pir*-116 electrocompetent E. coli to generate strains MYAS491 and 492, respectively. From these strains, I was able to prepare larger quantities of the suicide plasmids. Primers Tm727 and Tm729 were used to amplify *eptA-tnpR135-lacZ* from MYAS391. This cassette, along with puC18R6K-mini-Tn7T-Gm, were digested with BamHI and XhoI restriction enzymes to generate unidirectional compatible ends for intermolecular ligation. After ligation, this new plasmid was co-electroporated into MYAS382 (the SIVET strain) along with the helper plasmid, pTNS2, in order to generate MYAS510. MYAS510 is the complete reporter strain for the induction of PEtN modifications via *eptA*. The same procedure was used to generate the *arnT* reporter strain by extracting the *arnT-tnpR135-lacZ* cassette from MYAS394 with primers Tm728 and Tm729. This reporter strain was dubbed MYAS590.

Methods

Media – Cells were grown in LB unless otherwise stated. Both inducing and non-inducing media contained N-minimal media (5mM KCl, 7.5mM $(NH_4)_2SO_4$, 0.5mM K_2SO_4, 1mM KH_2PO_4, 0.1mM Tris-HCL, 0.2% glucose, 38mM glycerol, and 0.1% casamino acids in water). Inducing media contained N-minimal media at pH 5.8 with 10mM Mg^{2+} and 100µM Fe^{2+}. Non-inducing media contained N-minimal media at pH 7.7 with 10mM Mg^{2+}.

Polymerase Chain Reaction – All PCRs were performed using New England Biolabs® Q5® High-Fidelity DNA Polymerase. The protocol was followed according the NEB's standard protocol. Extension time was adjusted based on the length of the desired DNA with approximately 20 seconds per kilobase. The annealing temperature was adjusted based on the manufacturer's recommendation for primers (21). All future PCRs described in this thesis were similarly performed.

Red Recombinase Transformation with pKD46 – Bacteria containing the pKD46 plasmid to be transformed was grown overnight in LB at 30°C and subsequently diluted 1:100 in LB media. The diluted culture was grown at 30°C for 2 hours or until the OD reached approximately 0.3 to 0.4. L-arabinose (0.2%) was added and the culture was grown for approximately an hour or until the OD was approximately 1.0. The culture was spun down for 5 minutes at 3000 RCF and the supernatant was poured off. The cells were washed once with ice cold water and twice with ice cold glycerol (10%). The cells were centrifuged again and all supernatant was removed. The cells were resuspended in 70µL of glycerol (10%) and mixed thoroughly via pipette. 3-5µL of the amplified DNA insert was added and the cells were electroporated at 1800V and 300Ω. Afterward,

the cells were rescued with 1mL LB media and grown for 1-3 hours at 37°C. The cells were then plated on appropriate selection media and incubated overnight at 37°C.

Ligation – All DNA ligations were performed with New England Biolabs® T4 DNA Ligase according to their protocol and specifications. Molar ratios of DNA were calculated using NEB's NEBioCalculator™ (22).

Co-electroporation - 100µL of an overnight MYAS382 culture grown in LB media with tetracycline (5µg/mL) was inoculated into 10mL of LB media with tetracycline (5µg/mL) for 4 hours. The 10mL culture was centrifuged for 5 minutes at 3000 RCF and the supernatant was poured off. The pellet was washed with 2mL of a 10% glycerol solution three times. The pellet was then reconstituted in 90µL of 10% glycerol. 50ng each of the purified pTNS2 and puC18R6K-mini-Tn7T-Gm (containing either the *eptA* or *arnT* cassettes) plasmids was added and pipette-mixed. The cells were electroporated in a 2mm cuvette at 2500V and 200Ω. The cells were rescued with 1mL of LB media and incubated for 1.5 hours at 37°C. After incubation, the entire culture was plated on LB agar plates with gentamicin (10µg/mL) and HEPES buffer (20mM). After choosing colonies, they were substreaked on chloramphenicol, tetracycline, and carbenicillin in order to ensure that the cassettes had not cured and that the delivery plasmids were gone. The strain was confirmed by sequencing.

tnpR Allele Assay – Each of the pSEVA-containing strains (MYAS 436, 437, and 439) were grown overnight at 37°C in either inducing or non-inducing media, as described above. The OD of the cultures was taken and the cultures were then diluted by an appropriate factor for plating. 100 µL

of each of the six strains was then plated on either LB agar, LB agar with tetracycline (5 µg/mL) or LB agar with chloramphenicol (5 µg/mL).

Chapter 3

Reporter System Validation

Background

Now that the system has been constructed, it was employed to validate the initial theory that that the difference in modification level between K-12 and B was due to regulation of the *eptA* or *arnT* genes at the transcriptional level. Another possibility could have been that the lack of modification on K-12 was due to a strain-unique enzyme that facilitated removal after their addition. In order to debunk this, we used the β-galactosidase-producing gene *lacZ* that is also coupled to the promoters in the RIVET system. Strains that produce β-galactosidase will react with the substrate o-nitrophenyl-β-galactoside (ONPG) to produce a yellow color in proportion to the activity of *lacZ*. This provides a quantitative link between the activity of our promoter and the absorbance of the β-galactosidase reaction. In addition to MYAS510 and MYAS590, both within *E. coli* K-12, we generated two more strains by moving the RIVET-*lacZ* reporter cassette by P1vir transduction into *E. coli* B. Since the pUC18R6K-mini-Tn7T-Gm vector contained a gentamicin-resistance marker (GmR) that was also inserted during transposition, we were able to select for gentamicin in the transduction in order to move the RIVET cassette. The *eptA* cassette in *E. coli* B was dubbed MYAS591 and the *arnT* cassette in *E. coli* B was dubbed MYAS604. These four strains, MYAS510, 590, 591, and 604, were then grown in either LB as a control or one of two media conditions outlined in Chapter 2: inducing for LPS modification (dubbed “inducing”) or non-inducing for LPS modification (dubbed “non-inducing) (18). The expression level of the

promoters under each condition was then assessed quantitatively through the coupled *lacZ* gene via an adapted Miller assay (23, 24) (Figure 8).

Before the system could be used, we had to ensure both that it worked and that it operated at a sufficient sensitivity. To test this, MYAS510 was grown in both inducing and non-inducing media conditions. Each culture was then plated onto LB agar, LB agar with chloramphenicol, and LB agar with tetracycline in normalized amounts. The LB agar plates acted as a control to ensure that the total colonies on chloramphenicol and tetracycline are equal to the total before selection. Theoretically, under inducing conditions, the strains should lose tetracycline resistance and gain chloramphenicol resistance. Likewise, under non-inducing conditions, the strains should remain tetracycline resistant and not gain chloramphenicol resistance. Therefore, I determined the measure of curing—the result of the reporter undergoing activation to become chloramphenicol resistance—to be the ratio of chloramphenicol resistance colonies over the sum of tetracycline and chloramphenicol resistant colonies (Figure 10).

Results

The Miller assay shows full *lacZ* activity for both reporter cassettes in *E. coli* B irrespective of the media conditions. Conversely, both strains of *E. coli* K-12 only showed *lacZ* activity when grown in inducing media conditions. In all three media conditions and in each strain, there was roughly a fourfold increase in activity for the *arnT* cassette over *eptA*.

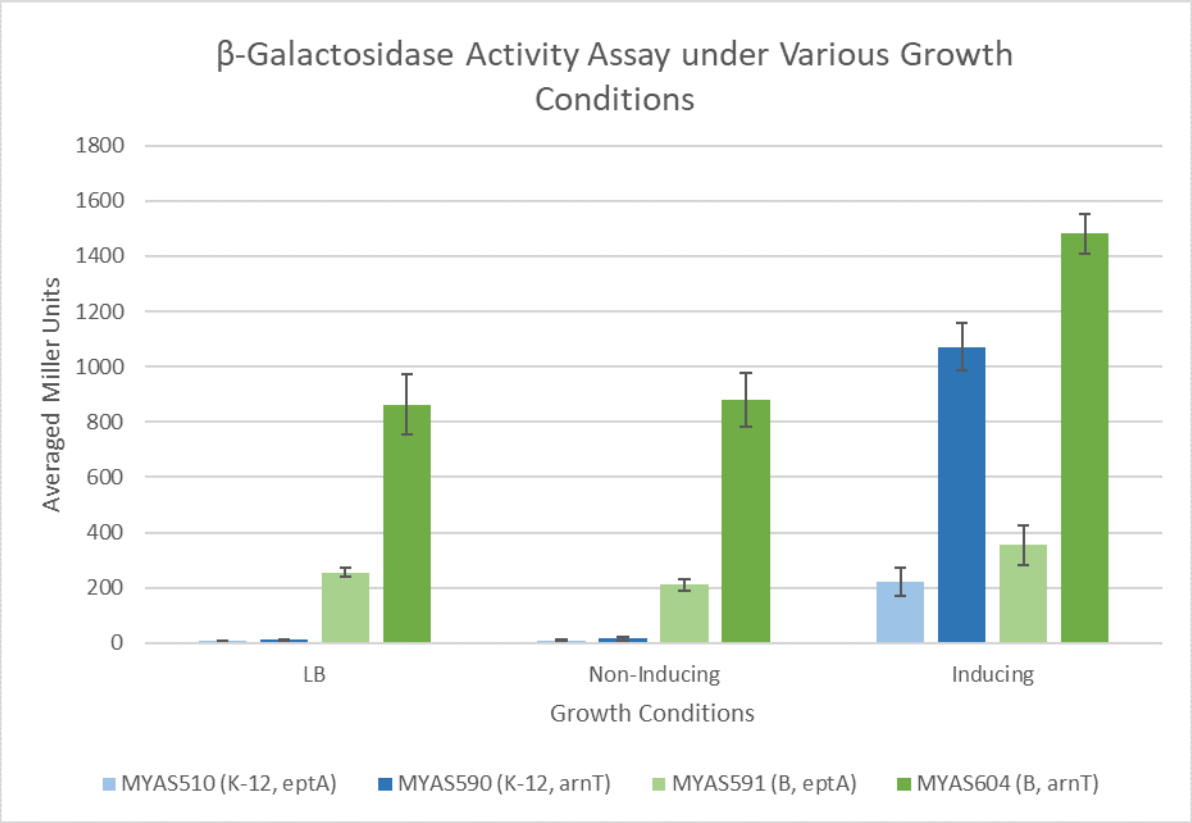


Figure 8: β-galactosidase activity assay.

The activity of the *eptA* or *arnT* promoter as determined by coupling with *lacZ* and assaying via β-galactosidase activity. The assay was performed on each of the two promoter cassettes in either *E. coli* K-12 or B grown in inducing or non-inducing media with LB as a control.

The curing assay demonstrated almost complete curing for both strains under inducing conditions and nearly no curing under non-inducing conditions (Figure 8).

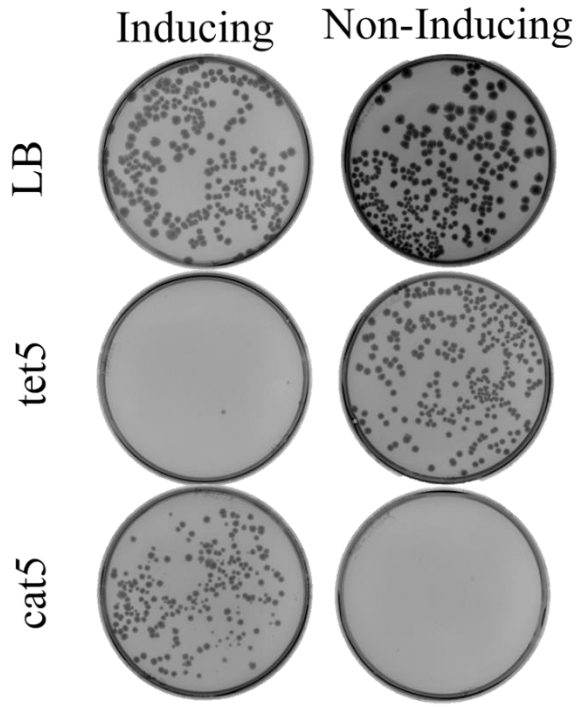


Figure 9: Curing assay example

The results from one trial of the curing assay for MYAS510. Results are representative of three independent experiments

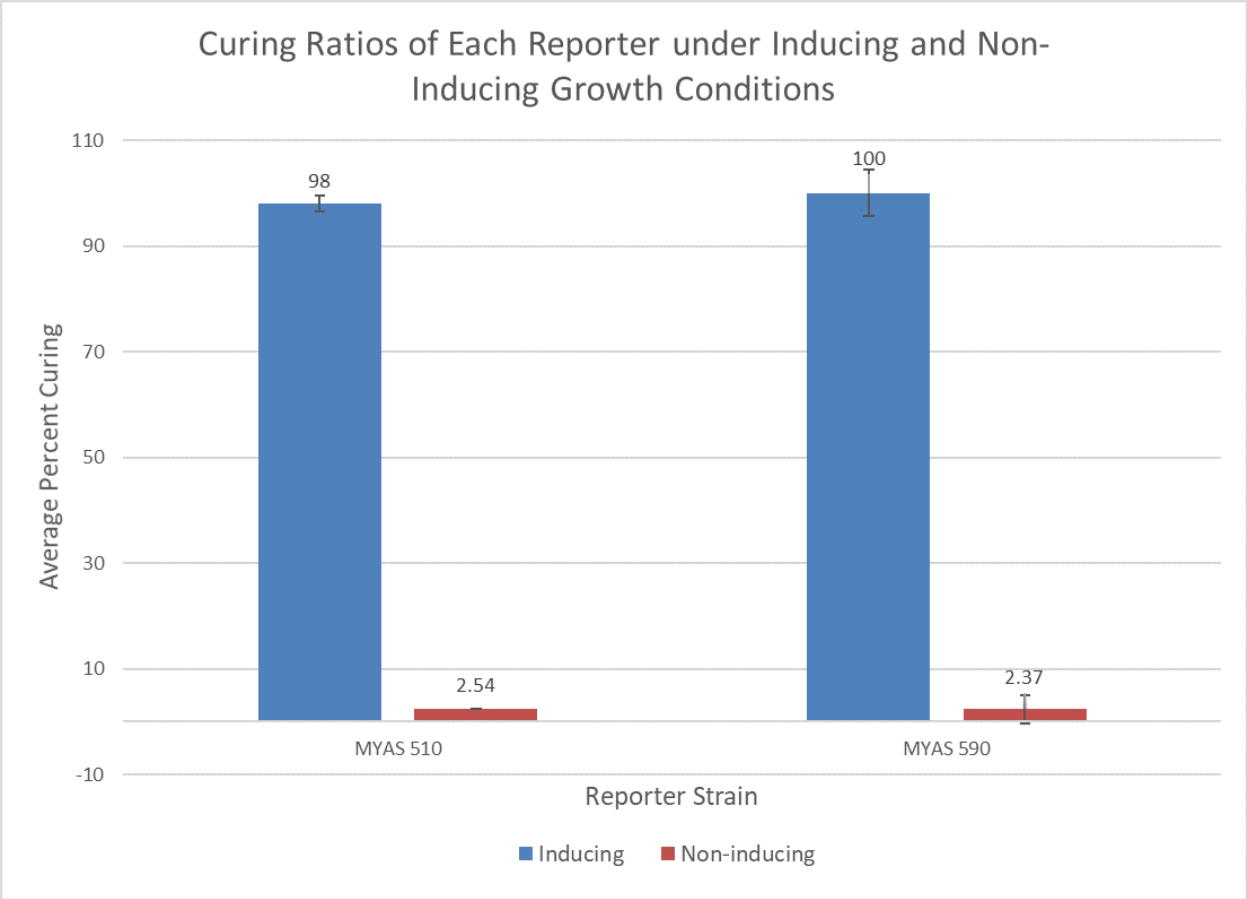


Figure 10: Curing ratio validation.

The results of the curing assay. Each strain was grown in inducing or non-inducing conditions and plated on LB as a control, tetracycline, and chloramphenicol. The results are averaged over three trials of each.

Discussion

When present in *E.* coli K-12, neither the *eptA* nor *arnT* promoter displays any activity when grown in LB or non-inducing conditions. These promoters only display activity when present in strictly inducing conditions. Conversely, when the promoters are present in *E. coli* B, they show constant activation, even under non-inducing conditions. These results demonstrate that there is a genetic basis for the differential expression of LPS modifications between *E. coli* K-12 and B that is unrelated to the promoter element itself. Whereas *eptA* and *arnT* seem to be constantly

suppressed in K-12, they are constitutively active within B (Figure 8). With this information, we could confidently move on to search for the activating alleles in B that are responsible for this effect.

The curing assay demonstrated that each of the two reporter strains possess very little spontaneous curing of the reporter cassette in the absence of any inducing factor. Ultimately, this bodes well for the level of background in the final experiment using the reporter system. Also, the reporter experiences nearly 100% curing in the presence of inducing factors. The data in Figure 10 demonstrate that the reporter system approximates a digital, binary system where induction can easily be differentiated from non-induction irrespective of inductive strength. It offers superiority over spectrophotometry-based assays such as GFP-coupling, for example. Using a promoter-GFP cassette to identify activation of modifications relies on a fluorescence spectrum that imparts a heavy background signal and relies on qualitative cut-offs to examine potential inducing conditions.

Methods

Media – Cells were grown in LB unless otherwise stated. Both inducing and non-inducing media contained N-minimal media (5mM KCl, 7.5mM $(NH_4)_2SO_4$, 0.5mM K_2SO_4, 1mM KH_2PO_4, 0.1mM Tris-HCL, 0.2% glucose, 38mM glycerol, and 0.1% casamino acids in water). Inducing media contained N-minimal media at pH 5.8 with 10mM Mg^{2+} and 100μM Fe^{2+}. Non-inducing media contained N-minimal media at pH 7.7 with 10mM Mg^{2+}.

Miller Assay – An overnight culture of the desired strain was grown in inducing and non-inducing media. The culture was diluted 1:50 into the fresh media of the same type and grown to an OD of 0.5-0.6. Each culture was centrifuged at 5000 RCF for 2 minutes and the supernatant was poured off. The pellets were then reconstituted in 2 mL A media (60mM K_2HPO_4, 33mM KH_2PO_4, 7.5mM $(NH_4)_2SO_4$, 1mM $MgSO_4 \cdot 7H_2O$, 3.4mM Citric Acid$\cdot H_2O$, pH 7 in water). The A media wash was repeated once more. The cultures were then serially diluted by factors of ten in order to provide a range of concentrations from which to obtain data. Cells were lysed by diluting 100µL of a given cell dilution with 900µL Z buffer (1mM $MgSO_4 \cdot 7H_2O$, 10mM KCl, 40mM $NaH_2PO_4 \cdot 2H_2O$, 60mM Na_2HPO_4, 50mM β-mercaptoethanol, pH 7 in water), 100µL chloroform, and 50µL 0.1% SDS. This reaction was performed in duplicate for each concentration in order to provide enough reaction mix for ten time-points. The reaction mix was vortexed vigorously for 10 seconds and then incubated for 10 minutes at 30°C. The aqueous layer was transferred to a new tube. 200µL of o-nitrophenyl-β-galactoside (ONPG) (4 mg/mL in A media) was added to each reaction mixture at the same time, the mixtures were briefly vortexed, and a timer was started. Before the first time-point at 5 minutes, 200µL of each reaction was aliquoted into rows of a 96-well plate. At each 5-minute interval, 83µL of Na_2CO_3 (1M) was added to the successive column to quench the reactions. The 96-well plate was then spectrophotometrically analyzed at absorbances of 420 nm and 550 nm. A graph of (Miller Units x Time) was plotted against time. Miller Units x Time was calculated as $1000 \times \frac{A420-(1.75\times A550)}{OD600\times 0.1}$. The slope of the linear portion of this graph was taken to be Miller Units. The Miller Units were then averaged for each time-point for the appropriate concentration of cells.

Lysate Preparation – An overnight culture of the donor strain was grown in LB at 37°C and then diluted one hundred-fold in 20mL LB (with 0.2% glucose, 5mM $CaCl_2$) in a baffled flask. The culture was grown for 1 hour at 37°C. Then, 50µL of P1vir bacteriophage was added. This was incubated for approximately 3 hours or until lysis occurred. The culture was then centrifuged at 3000 RCF for 10 minutes. The supernatant was filter sterilized and stored at 4°C.

Transduction – An overnight culture of the recipient strain, grown in LB at 37°C, was centrifuged at 4500 RCF for 2 minutes. The supernatant was discarded and the pellet was resuspended in a half volume of LB (with 10mM $CaCl_2$, 5mM $MgSO_4$). 100µL of the recipient cells were mixed with 100µL of serially diluted lysate or 100µL of the above LB as a control. This mixture was incubated in a 37°C water bath for 30 minutes. Then, 200µL of citric acid (1M) was added to each reaction mixture before combining the whole mixture with 1mL of LB media. The mixture was incubated for another hour at 37°C. After incubation, the cultures were pelleted in a centrifuge at 4500 RCF for 2 minutes, washed once with LB media, and then resuspended in 100µL LB media. The culture was plated on selective agar media and grown at 37°C.

Curing Assay – An overnight culture from a glycerol stock of the desired strain was grown in both inducing and non-inducing conditions. The cultures were then centrifuged at 4000 RCF for 2 minutes and then the supernatant was removed. The pellets were reconstituted in LB and grown for 1 hour at 37°C. The OD_{600} of each culture was taken. The inducing culture was diluted to normalize its OD_{600} to the non-inducing culture. Both cultures were then diluted by a factor of 10^6 and plated on LB agar, LB agar with chloramphenicol (5µg/mL), and LB agar with tetracycline (5µg/mL).

Chapter 4

Transposon Library Construction and Transduction

Background

A full genomic screen of *E. coli* B for activating alleles necessitates a method for full-coverage analysis of the genome. Transposons are an efficient and convenient method to thoroughly tag a genome with a selectable marker in a random fashion. By establishing linkage between all possible alleles and a selectable marker, we are able to sample this library for alleles that activate our reporter by transduction (Figure 5). NGS can be used to scan the genome for the linkages between *E. coli* B DNA and the transposon. After NGS, the identity of these inducing alleles can then be elucidated by locating the transposon and mapping the linked DNA back to *E. coli* B. Transposon mutagenesis coupled with whole-genome sequencing is a powerful tool for regulatory element sequencing in any number of microbiological application. Commercial transposon insertion kits are available to purchase; however, for our purposes, we worried that they would not provide the genome coverage that was necessary to screen for allelic differences and not just entire open reading frames. We employed a mariner-based transposon system which affords advantages over other transposons because of its low site-specificity. It only requires a TA dinucleotide for insertion, allowing us to obtain high coverage in the B genome (25). The system that we employed was adapted from a system designed in *Staphylococcus aureus* using the HMAR transposon and transposase (25). The principle is to separate the mobile element from the transposase and place them each on conditionally-replicative plasmids. The transposon-containing

plasmid contains a *pir*-dependent origin of replication and is grown in a *pir*$^{+}$ strain in order to recover purified plasmid. Meanwhile, the transposase-containing plasmid contains a temperature-sensitive origin of replication that only permits growth at 30°C instead of 37°C. The transposase gene itself is also under the control of the arabinose-inducible *araBAD* promoter. The transposase plasmid is electroporated into *pir*$^{-}$ *E. coli* B and grown without arabinose at 30°C. The transposon plasmid is electroporated into this strain which is then grown at 37°C with arabinose. The transposase plasmid expresses transposase so that the transposon cassette can move into the genome. Antibiotic resistance to kanamycin will only be obtained if the transposon hops from the plasmid into the chromosome; otherwise, both plasmids fail to replicate and are lost. This system is designed so that only one transposon is inserted per bacterium. It also limits the amount of transposase expressed as it is cytotoxic.

Random 90 kilobase portions of the *E. coli* B library was then to be transduced into our reporter strains, MYAS510 and MYAS590, via P1*vir*. This new library was then selected against kanamycin to select for strains of the reporter that received a transposon. After repeating the transduction and selection to ensure high-density coverage of the transposon, we further subjected this pool to chloramphenicol selection to enrich for strains that both received a transposon and activated the RIVET-SIVET reporter system.

Results

A Q5 polymerase PCR was performed to amplify the C9 HMAR transposase from the pWV01 plasmid with primers TM903 and TM904 to add extensions containing ligation compatible restriction sites. Purified pKD46 plasmid containing a temperature-sensitive origin of

replication was used as the template in another Q5 polymerase PCR to amplify the vector backbone with primers TM909 and TM910. The C9 HMAR transposase was then cloned into the pKD46 vector with In-Fusion cloning and then transformed into commercially competent *E. coli* (Figure 11). The plasmid was recovered and digested to check for cloning accuracy. After confirmation, the plasmid containing strain was stocked as MYAS654. This strain must be grown at 30°C in order to propagate the plasmid.

Next, we constructed the transposon-containing (Tn) suicide plasmid. We engineered this plasmid with a cosmid packaging site (*cos*) and a conjugation site (*oriT*) so that we would have other delivery options in case electroporation yielded too low of an efficiency. However, before we could generate the plasmid, we first had to insert the kanamycin-resistance marker (*kanR*) into the transposon cassette itself. *KanR* was amplified by Q5 polymerase PCR from pKD4 with

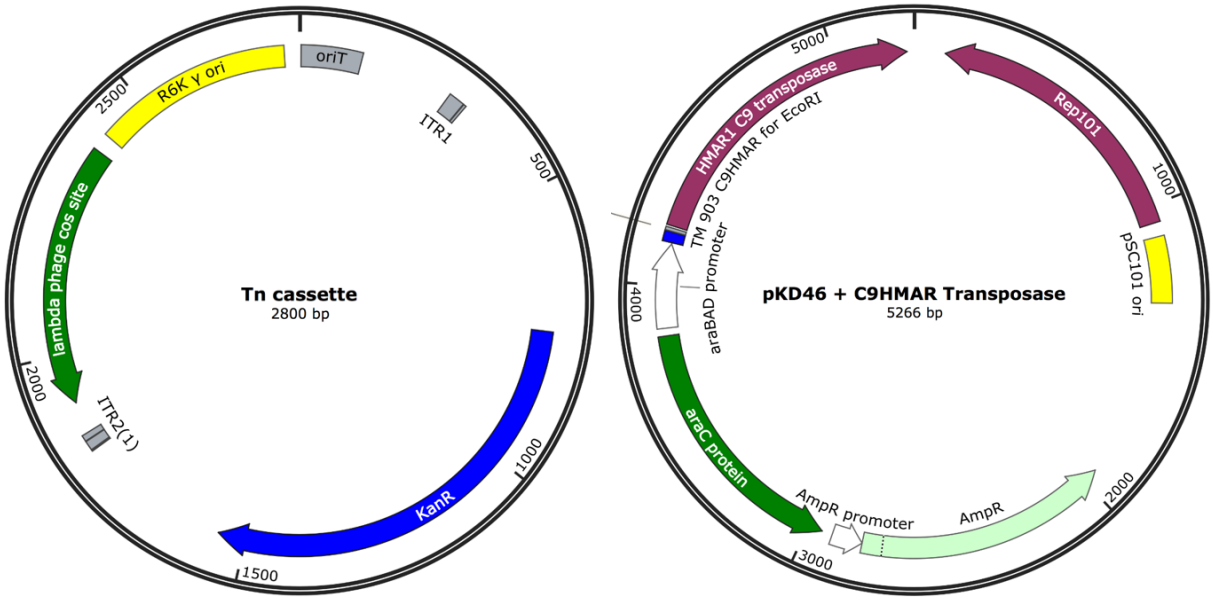


Figure 11: Plasmid maps of transposon library generation system

The plasmid maps showing the relative genetic components of the Tn cassette plasmid (left) and the temperature-sensitive C9 HMAR transposase plasmid (right).

TM916 and TM917. This fragment was then inserted into digested Tn-containing vector pTM223 via In-Fusion and heat shock transformed into *E. coli*. This new pTM223 with *kanR* plasmid was purified and further used as a template for PCR. To generate the final Tn plasmid, four different Q5 polymerase PCR reactions were performed with different templates and primer pairs to generate four fragments for ligation. The component containing the *cos* site was amplified from the pWEB vector with primers TM294 and TM295. The transposon cassette with *KanR* was amplified from our pTM223+*kanR* plasmid with TM290 and TM291. The *pir*-dependent origin of replication was amplified from pKD3 with TM931 and TM932. Finally, the *oriT* was amplified from pEX18 with TM933 and TM934. Each piece was designed with overlap in order to ligate them together; however, the four-piece ligation by In-Fusion proved difficult. Instead, we performed two overlap-extension PCRs to fuse two pieces together at a time. First, the *oriT* and Tn cassette segments were joined with primers TM933 and TM291. Next, the *cos* and *pir*-dependent *ori* segments were joined with TM295 and TM932. The two larger pieces were then joined to create the final plasmid construct by In-Fusion before heat shock transformation into *E. coli* (Figure 11). After confirming the plasmid's composition by digestion and sequencing, the strain containing the plasmid was stocked as MYAS664.

The temperature-sensitive plasmid containing the transposase, C9 HMAR, under the control of the arabinose-inducible promoter was cloned into *E. coli* B and dubbed MYAS 667. This strain is normally grown under the permissive temperature of 30°C. Then, the purified Tn plasmid was electroporated into MYAS667, a *pir*- strain, which was grown with 0.2% L-arabinose and under non-permissive temperature. Under these conditions, only one copy of the Tn cassette would be present in each cell. The transposase would be activated but the plasmid would not be able to replicate. This ensures that transposase production is limited due to its cytotoxicity. The

process of electroporation was repeated eight times on large plates in order to generate a large transposon library. The pooled library was estimated to contain 582,400 transposants. A negative control was also acquired in this manner by stocking a single colony after the transduction. The probability that the transposon carried an activating allele into the single colony that was chosen is extremely low, allowing this to act as a negative control. Finally, the transposon library and low-diversity control library were packaged in P1vir phage in order to transduce into our SIVET/RIVET reporter.

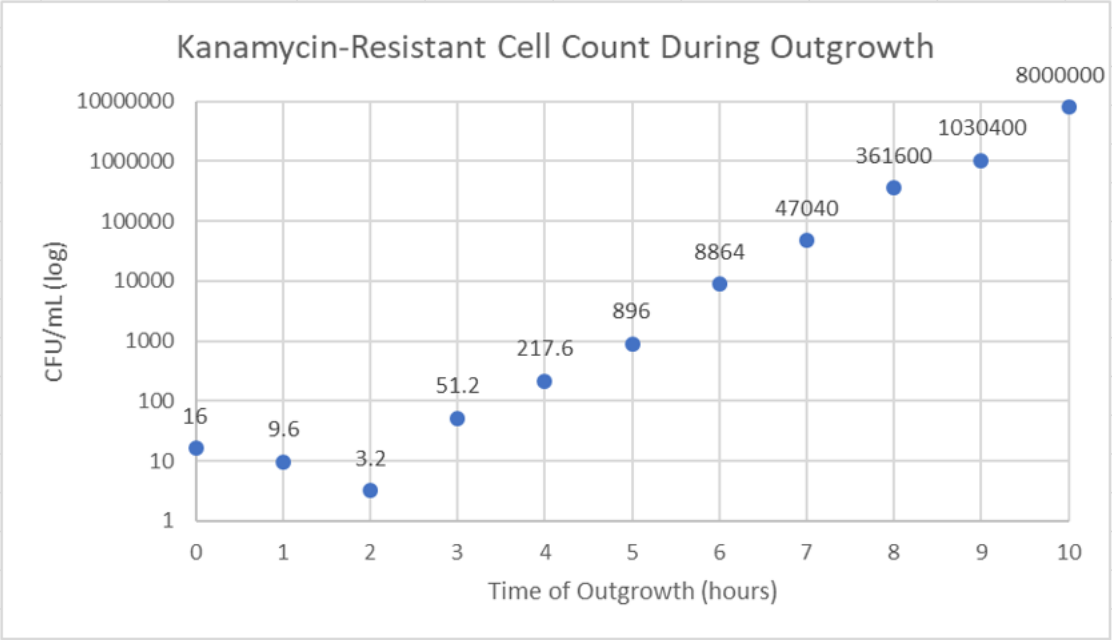


Figure 12: Kanamycin-resistant cell count during outgrowth

Hourly time-points were taken from the outgrowth culture of MYAS510 after the transduction. Successful transductants would receive a transposon containing a kanamycin-resistance marker along with genomic DNA from *E. coli* B.

The liquid-liquid P1vir transduction was used to move 90kb segments of the transposon library from *E. coli* B to our reporter strain (26). Eight transduction reactions were performed in order to move the B transposon library into MYAS510. We were aiming for approximately 100,000 unique transductants. Additionally, a transduction was performed with the single colony

library as a control. During the outgrowth period, hourly time-points were extracted from one experimental transduction and the control transduction. These time-point extractions were plated on LB with kanamaycin (30 µg/mL) or chloramphenicol (10 µg/mL) in order to obtain cell counts.

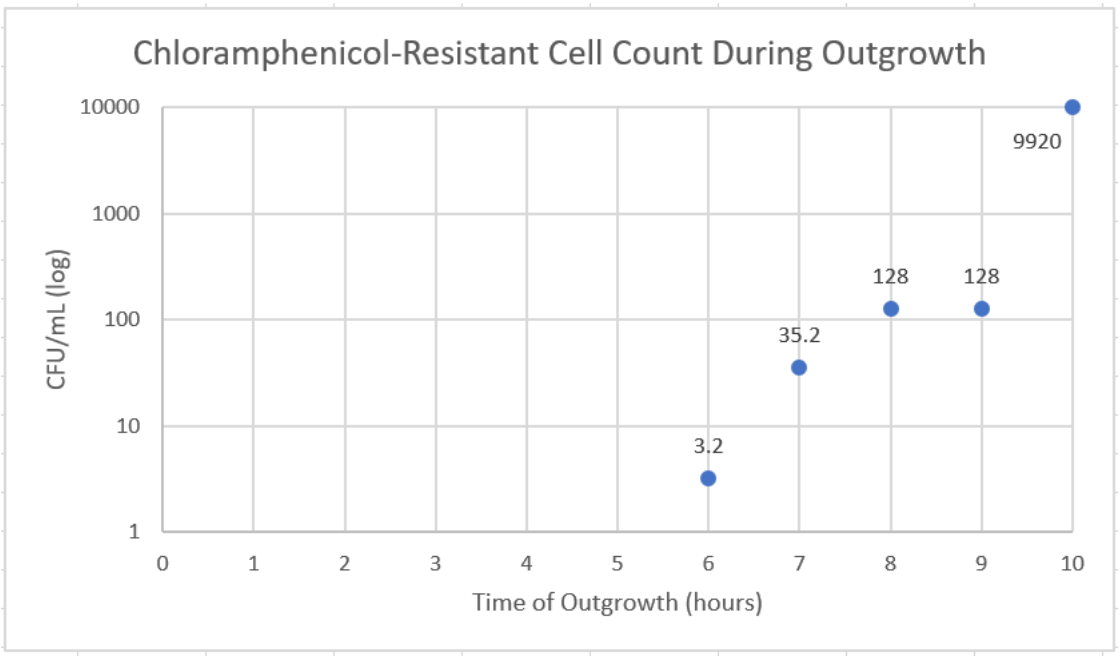


Figure 13: Chloramphenicol-resistant cell count during outgrowth

The hourly time-points taken from the outgrowth culture of MAS510 were also grown on chloramphenicol to assess the level of induction of the reporter.

We were checking to ensure that there was no population bottleneck during the outgrowth that could decrease the diversity of transposon insertion. There was a slight dip in kanamycin-resistant colonies at two hours; however, this change was relatively small compared to the initial cell count. After hour two, the cell count rose steadily according to expectation (Figure 12). We do not believe that this dip represents a significant loss of diversity in the reporter library. We repeated this cell count for chloramphenicol resistance to initially assess the number of activated reporters in the pool of transductants as a function of time. Aside from the apparent stagnation in growth at hour nine, chloramphenicol colonies appeared at hour six and steadily grew until the end of the

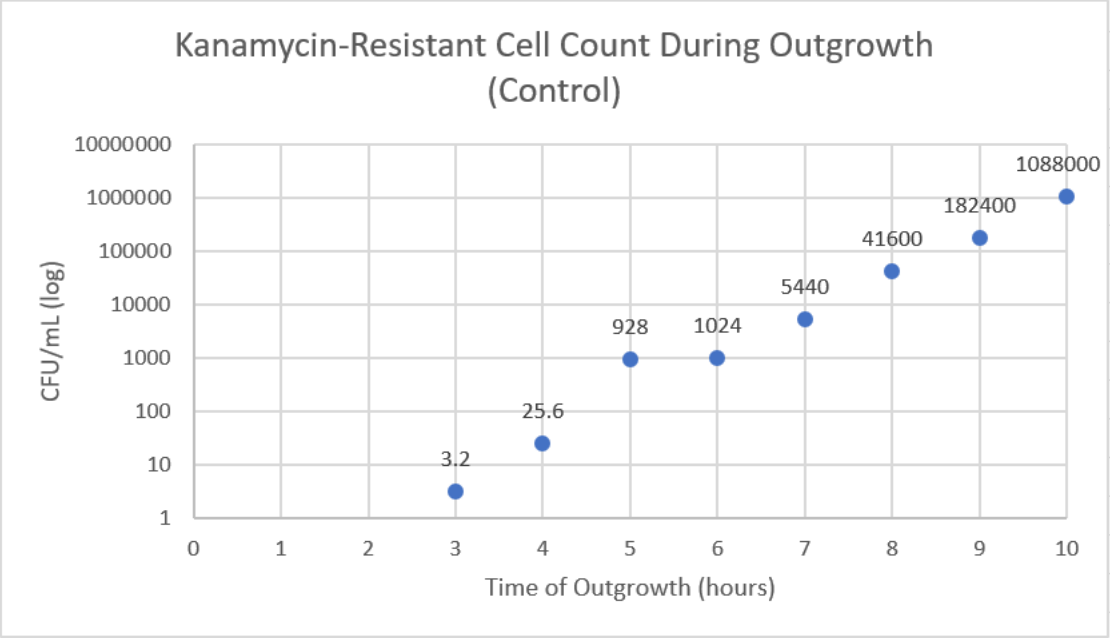


Figure 14: Transduction control cell count during outgrowth

The hourly time-points taken from the control transduction where the single colony library donor lysate was used with MYAS510 as the recipient.

outgrowth (Figure 13). Finally, this time-point process was repeated for the control transduction where the donor library lysate originated from a single colony. This data also fall within expectation as there is steady growth throughout the outgrowth period (Figure 14). The control time-points were also tested for chloramphenicol resistance; however, there were found to be no chloramphenicol-resistant colonies in the single colony library transduction throughout the course of the experiment. This suggests that the appearance of chloramphenicol resistance in the experimental, diverse library transduction is due to a transferred activating allele and not to spontaneous curing of the reporter.

After outgrowth, the cultures were concentrated and stocked. Final cell counts were obtained for each trial after outgrowth was completed. As described in the methods, cell counts were obtained for the transductions immediately after phage incubation and before the outgrowth

period. These cell counts theoretically represent the number of unique transductants from each trial.

Now that we had a library of our *E. coli* K-12 RIVET-SIVET *eptA* reporter containing random 90 kbp genomic DNA fragments from *E. coli* B, we had to enrich the library for activated reporters. These trials must be pooled proportionally to maintain equal representation from the unique transductants in each trial. To calculate this, we multiplied the unique transductant count by 100. This theoretically translates to 100 reads of each unique transductant. That count was then divided by the total KanR cell count in the trial to obtain the proportional contribution to the pool. The final pooled, transduced reporters were plated on ten large LB plates with chloramphenicol (10 µg/mL) immediately after combination to select for *eptA*-promoter activated reporters. After growing overnight, a small amount of LB was pipetted on each plate and all of the colonies were scraped and pooled. The pooled library was aliquoted and stored as the final samples.

Table 2: Final transduction cell count data in MYAS510

Transductant Trial	Unique Transductant Count	Final KanR Cell Count (CFU/mL)	Final CatR Cell Count (CFU/mL)	CatR/KanR Ratio	Contribution to Pool (mL)
1	30000	6.08E+09	2.81E+07	0.46%	0.49
2	8800	1.36E+09	6.60E+06	0.49%	0.65
3	9200	3.02E+09	1.55E+07	0.51%	0.30
4	9600	1.81E+09	1.52E+07	0.84%	0.53
5	9600	3.67E+09	2.43E+07	0.66%	0.26
6	6400	4.00E+08	4.00E+05	0.10%	1.60
7	10000	1.00E+09	2.30E+06	0.23%	1.00
8	13200	1.35E+09	8.90E+06	0.66%	0.98
Total	96800				5.81

Discussion

The plasmids used for transposon library generation were thoroughly sequenced in order to confirm the success of their cloning. Though the density and coverage of the transposon library in *E. coli* B can only quantitatively be assessed after next-generation sequencing, plating allowed us to initially assess the coverage that we could expect.

The time-point data in Figures 12-14 allowed us to conclude that the transduction did not significantly bottleneck the diversity of the transposon library. Additionally, it allowed us to confirm before sequencing that the transposon library is capable of activating the reporter and inducing chloramphenicol resistance. It is also important to note that the control library demonstrated a lack of spontaneous curing of the reporter. The total number of unique transductants is consistent with our goal. Additionally, the ratio of chloramphenicol-resistant colonies to total kanamycin-resistant colonies implies good representation of activated reporters in the library of transduced reporters (Table 2). Each trial was pooled in order to perform the genomic DNA extraction for NGS.

Unfortunately, we were not able to complete the final transduction experiment with MYAS590, the *arnT* reporter strain, as the recipient. We inexplicably and repeatedly obtained a marked reduction in transduction efficiency and were not able to meet the same unique transductant target as with MYAS510. We moved forward with the transduced samples of MYAS510 to perform genomic DNA preparation and NGS.

Methods

In-Fusion Cloning – The procedure was followed from the Takara Bio USA In-Fusion® HD Cloning Kit User Manual (27).

Transposon Library Construction – MYAS667 was grown overnight in LB media with carbenicillin (100 µg/mL) at 30°C. The culture was diluted 1:100 and then grown for 4 hours in LB with carbenicillin (100 µg/mL) and L-arabinose (0.2%) until it reached an OD of 1. The culture was then electroporated with approximately 1 µg of tn-cassette plasmid at 1800V and 250Ω. The electroporation was rescued with 1mL of LB media and grown for 1.5 hours at 37°C. The culture was diluted 1:100 to plate onto LB agar with kanamycin (30 µg/mL) in order to get a cell count of unique transductants. The rest of the outgrowth was plated onto large kanamycin plates in order to collect a large library.

Donor Library Lysate Preparation – The donor *E. coli* B transposon library was inoculated into LB at an OD of approximately 0.05-0.1. The LB media was then supplemented with 0.1% glucose. The donor strain was grown to an OD of 0.5 before the media was supplemented with $CaCl_2$ (5mM). To the donor strain, 50µL of P1vir phage was added and the bacteria were incubated without shaking for 20 minutes. Top agar was added to the media and the mixture was readjusted to 0.1% glucose and 2.5mM $CaCl_2$. The mixture was then poured over LB plates (0.1% glucose and 2.5mM $CaCl_2$). The plates were incubated overnight at 37°C. 2mL of LB media was then added to the top of each plate and the top agar was scraped into a centrifuge tube. The top agar was centrifuged at 14.5 kGs for 10 minutes. The supernatant was filter sterilized and stored as the donor lysate.

Liquid-Liquid P1vir Transduction – An overnight culture of the reporter was grown in LB with tetracycline (5 µg/mL) and gentamicin (10 µg/mL). In the morning, the culture was diluted 1:100 into 100 mL of LB with tetracycline (5 µg/mL) to an OD of 1. This culture was centrifuged at 6000 RCF for 10 minutes. The pellet was resuspended in 10 mL of transduction buffer (LB with 5mM $CaCl_2$ and 10mM $MgSO_4$). 5 mL of donor lysate was then added and the reaction was incubated at 37°C for 0.5 hours without shaking. The culture was recovered with 25 mL of LB with sodium citrate (10mM). The culture was then incubated again at 37°C for 0.5 hours without shaking. The culture was spun down again at 6000 RCF for 10 minutes. The pellet was resuspended in 4 mL sodium citrate (1 M). From this resuspension, 10 µL was taken and combined with 90 µL phage buffer. This was plated on LB agar with kanamycin (30 µg/mL). The count of unique transductants was calculated as $\frac{\#colonies}{0.01\ mL}(4\ mL)$. The 4 mL culture was inoculated into a 250 mL culture of LB with kanamycin (30 µg/mL). This was grown for 10 hours at 37°C with shaking to allow for integration of the transposon library and activation of the reporter system. The culture was spun down at 6000 RCF for 10 minutes and the pellet was resuspended in 10 mL LB with glycerol (30%). 1 mL aliquots of this resuspension were stocked. This protocol can be scaled up to conduct multiple trials of transduction simultaneously.

Chapter 5

Next-Generation Sequencing

Background

Next-generation sequencing (NGS) allows for the efficient screening of entire genomes for open reading frames, regulatory elements, and other pieces of genetically-encoded information. We employed NGS for the purposes of transposon sequencing (Tn-seq), that is, to screen for activating alleles in the genome that have been flagged with a randomly-inserting transposable element. We used Illumina®, a sequence-by-synthesis method, in order to sequence our transposon libraries (28). In total, we prepped three samples of genomic DNA for sequencing. If a transposon hopped close to a *eptA*-promoter activating allele, then this particular transposon should be overrepresented in the NGS data. In theory, a bell-curved enrichment of approximately 90 kb should span the position of maximal transposon insertion due to the phage packaging size and the linkage between strain B gDNA and the transposon cassette. First, we sequenced the pooled transposon library in *E. coli* B. This is to assess the coverage of our library and look for any holes in the genome or bias in the transposase. Second, we sequenced the pooled stock of transduced reporters in *E. coli* K-12 that were unchallenged with chloramphenicol in order to establish a baseline of transposon insertion into the reporter. Thirdly, our collection of chloramphenicol-resistant, activated reporters was sequenced in order to assess hotspots of transposon insertion.

Results

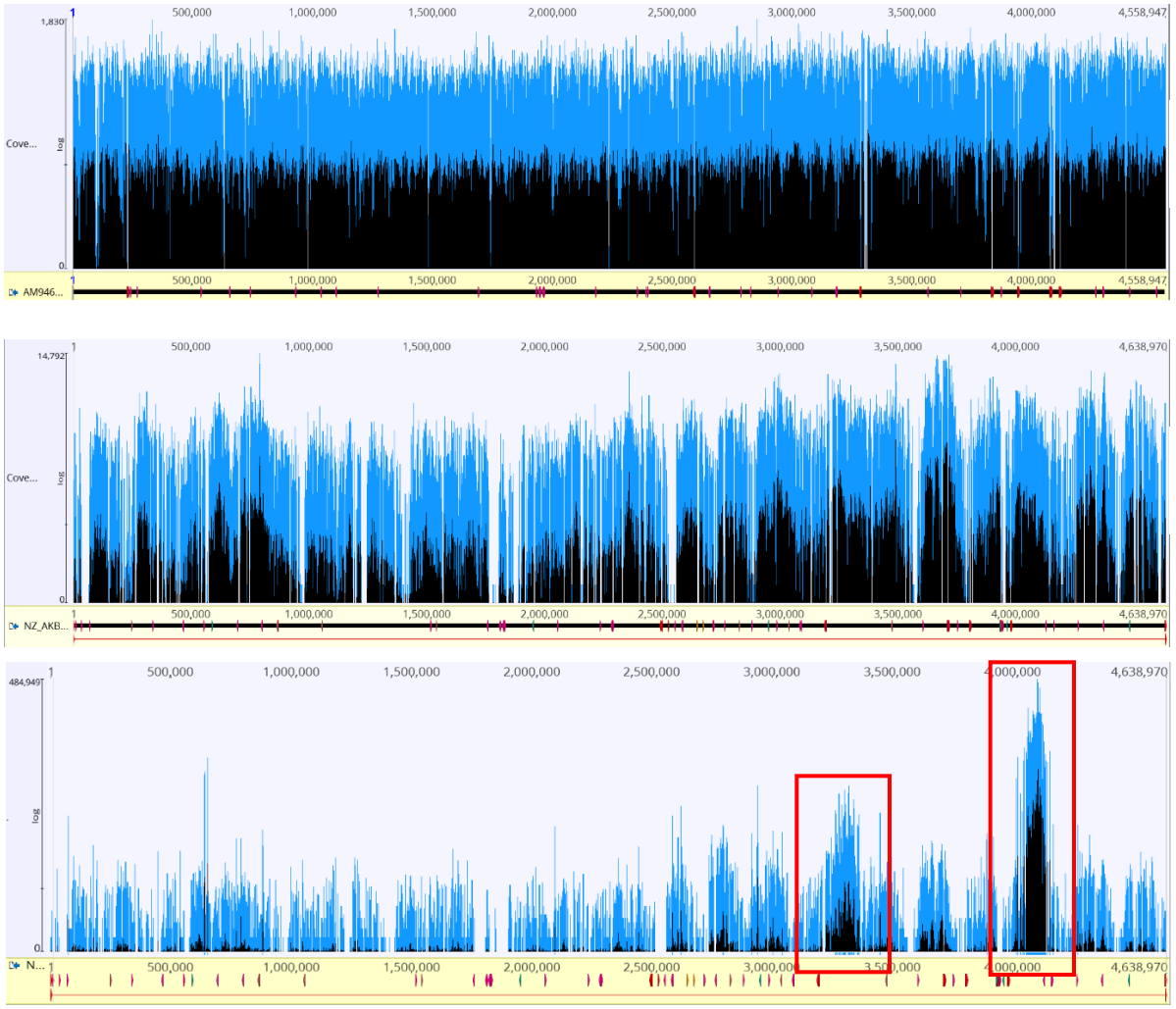


Figure 15: Next-generation sequencing alignments

The three panels above represent the sequence coverage and alignments to the given reference genome for the three gDNA samples. The height of the bars represents the read number with blue representing the maximum reads and black representing the mean reads. A) The sequence alignment of the transposon library to the *E. coli* B reference genome (Accession: AM946981). B) The sequence alignment of the transposon library inserted into the reporter strain by transduction and not selected for activated reporters. This was aligned to the *E. coli* K-12 reference genome (Accession: NZ_AKBV01000001). C) The sequence alignment of the transposon library inserted into the reporter strain and selected for activated reporters on chloramphenicol. This was also aligned to the *E. coli* K-12 reference genome (Accession: NZ_AKBV01000001). The two boxed peaks represent the genome clusters with the highest read coverage as denoted by the blue hash marks underneath. The center of each peak is the genomic position of the putative activating allele since there was the highest degree of linkage with the transposon cassette.

The three genomic DNA samples were successfully purified and sequenced. The sequences were then trimmed and mapped back on to the appropriate *E. coli* reference genome. Linkage between the *kanR* cassette and the strain B gDNA and transduction in 90 kb segments

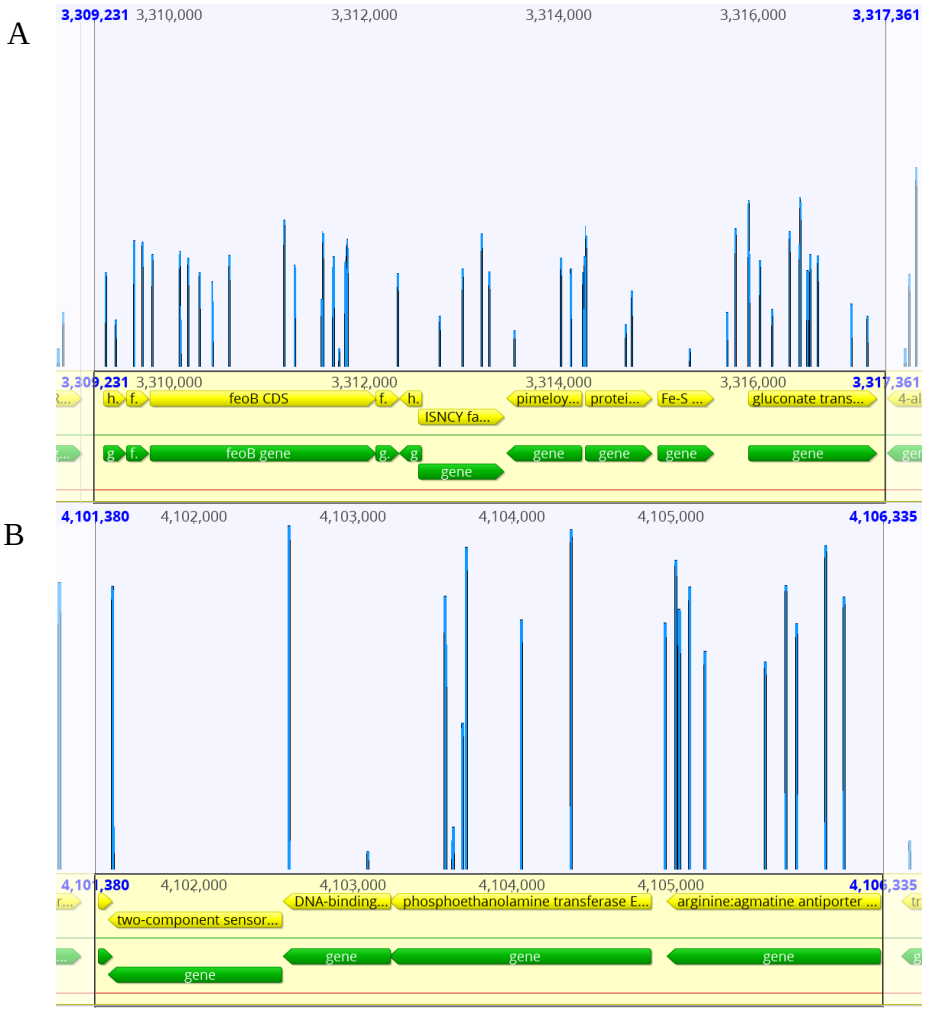


Figure 16: Significant transposon insertion sites

A) The magnified view of the smaller significant hotspot from Figure 15C containing approximately 5 kb centered around the maximal peak. B) A magnified view of the larger significant peak from Figure 15C containing approximately 5 kb centered around the maximal peak.

results in a bell-curve distribution of approximately 90 kb centered around a peak. After identifying the two most significant peaks at the approximate positions 3,313,000 and 4,103,000

of the genome, we selected a roughly 5 kb range spanning the maximal read position for further analysis (Figure 16). The genes in this range represent the highest probability of linkage to the allele that activated the reporter. The smaller of the two significant hotspots contains, among others, genes for iron transport and utilization as well as gluconate transport. The second, larger hotspot notably contains phosphoethanolamine transferase, *eptA*, and the two-component sensor *basSR*.

Discussion

The genomic DNA preparations and subsequent NGS of the three samples was successful. Our in-house transposon library generation platform in *E. coli* B as described in Chapter 4 resulted in an extremely dense, high coverage library with no significant gaps or biases (Figure 15A). When this library was transferred via P1vir transduction to MYAS510 and selected for the presence of the transposon alone, there was uniform transduction insertion of the transposon and B DNA. There are a few significant gaps which can be attributed to strain B specific prophages and genomic islands, such as those for varying sugar utilization, that have no homology in the strain K-12 genome (Figure 15B). Therefore, there was inefficient transduction of the transposon-linked strain B DNA into those areas in strain K-12. When the reporter containing the transposon/B DNA library was selected against chloramphenicol to enrich for activated reporters, two significant hotspots are immediately noticeable centered around positions 3,313,000 and 4,103,000 of the genome (Figure 16). We chose these loci for further analysis.

In our original assessment of *eptA*'s genetic regulation, no differences in the promoter region between *E. coli* K-12 and B were noted upon alignment. However, an alignment of the

whole gene between the two strains using BLAST (29) reveals nine SNPs in the flanking sequence (Table 3). Interestingly, there are five SNPs in the *basSR* system. As stated in Chapter 1, the products of these genes sense, among other signals, iron in order to modulate the expression of *eptA*, *arnT*, and *lpxT*, all of which modify LPS (13). If a gain-of-function constitutive *basSR* allele in *E. coli* B results in overactivation of *basSR*, then *eptA* would be constitutively active and *E. coli* B would present a highly LPS-modified phenotype. There were virtually no transposon insertions within the *basSR* system, only around it (Figure 16B). This strongly supports that a gain-of-function allele is present in *E. coli* B that was transferred to elicit activation of our reporter. In fact, there are known mutations that activate *basR* and the highly homologous *pmrA* gene in *Salmonella typhimurium* (3, 30, 31). While none of the *E. coli* B SNPs are the same, Froelich et al. has characterized other mutations, such as *basR*G53V, which result in increased modification with L-Ara4N via *arnT* and subsequent polymyxin resistance (30). Hence, there is precedent for a SNP to activate *basSR*.

The smaller hotspot contained several more genes of interest that we hypothesize could lead to increased expression of *eptA*. It is well documented that the modifications conferred to LPS by *eptA* and *arnT* reduce the bacterium's susceptibility to metal poisoning by reducing the negative charge on LPS (11, 13). Hence, the *basSR* system engages *eptA* and *arnT* when there is a high iron concentration. The *feoABC* iron transporter operon was located squarely in the maximal peak of the smaller hotspot. There are nine SNPs between *E. coli* K-12 and B within the largest protein of the *feo* operon, *feoB* (Table 3). If the *E. coli* B allele expresses lower levels or a less active version of the iron transporter than *E. coli* K-12's, then B will have a higher periplasmic iron concentration that will be sensed by *basS*. This would potentially result in upregulation of *eptA*. In contrast to the other hotspot, the *feoABC* operon has a high frequency of transposon insertions, indicating that

either a knockout of *feoABC* or a highly under-functioning allele is responsible for reporter activation (Figure 16A).

Table 3: Putative causative alleles of differential phenotype

	Gene	Function	Relative SNPs
Fig. 16A	***feoA***	Ferrous iron transporter (32)	0
	feoB		9
	feoC		7
	yhgA	Putative transposase (33)	12
	bioH	Pimeloyl-ACP methyl ester carboxylesterase (34)	2
	gntX	Gluconate periplasmic binding protein (35)	6
	nfuA	Iron/Sulfur protein biogenesis (36)	10
	gntT	High-affinity gluconate transporter (37)	2
Fig. 16B	***basS***	Transmembrane sensor (13)	1
	basR	Transcriptional regulator (13)	4
	eptA	Phosphoethanolamine transferase (13)	9
	adiC	Arginine:agmatine antiporter (38)	0

Methods

Genomic DNA Preparation – Approximately 2-4 mL of *E. coli* cells at an OD of 1-2 were pelleted and resuspended in 100 µL P1 buffer with RNase. 500 µL of GES buffer (5.1M guanidinium thiocyanate, 100mM EDTA pH 8, heated at 65°C in 20 mL ddH_2O to dissolve and cooled to room temperature, 0.5% sarkosyl, filtered through a 0.45 µm filter) was added and the mixture was vortexed. 250 µL of cold ammonium acetate (7.5M) was added and the mixture was incubated on ice for 10 minutes. 500 µL of chloroform was added, vortexed, and centrifuged at maximum speed for 10 minutes. The supernatant was removed to Lobind microcentrifuge tubes and combined with 400 µL isopropanol. The tubes were inverted to mix until cotton-like DNA threads were visible. The tubes were centrifuged at max speed for 5 minutes in order to pellet the mixture. It was washed

with ethanol (70%) twice with 2-minute spins in between. The final pellet was dried in a biosafety cabinet for approximately 15 minutes. The pellet was resuspended in 100 µL elution buffer (0.5x), vortexed, and incubated with shaking at 37°C for 1 hour. The concentration was measured and the gDNA was stored at -80°C.

Illumina Sample Preparation – NotI Digestion – 10 µg of gDNA was resuspended in 100 µL of autoclaved ddH_2O. To this DNA, the following reagents were added for a NotI digestion reaction: 60 µL Buffer 3 (10x), 3 µL BSA (200x, 20mg/mL), 2 µL NotI, and 435 µL ddH_2O. This reaction mixture was vortexed gently, inverted, and spun down briefly. The reaction was incubated at 37°C for 7 hours, mixed halfway through, and then incubated at 70°C for 20 minutes. After cooling to room temperature in a water bath, the digests were stored at -80°C.

Illumina Sample Preparation – Selective Precipitation of DNA – To the thawed gDNA, 200 µL of 4x Buffer (4.4M NaCl, 1M Na_2PO_4 pH 7.5, 32% PEG8000, in autoclaved ddH_2O) was added and vortexed to mix. This was incubated in an ice water bath for 12-16 hours. Afterward, the mixture was centrifuged at maximum speed at 4°C for 20 minutes. The large gDNA containing pellet was washed once with 1 mL of a 1x dilution of the 4x Buffer and then centrifuged again at maximum speed at 4°C for 10 minutes. It was then washed twice with 1 mL of 70% ethanol with room temperature spins (18000 RCF) for 5 minutes in between. The pellet was dried for 15 minutes in the biosafety cabinet before being dissolved with 50 µL of 0.5x elution buffer. The dissolved DNA was then stored at -80°C.

Illumina Sample Preparation – Ligation of Adaptors to Digest – Illumina adaptors TM214 and TM215 (Table 6) were resuspended in elution buffer at a concentration of 100 µM. Equal volumes of each adaptor were combined with one tenth volume NaCl (1M) and annealed in a thermocycler (95°C for 5 minutes, cooled to 4°C with ramp-down at 0.1°C/s). The adaptors were diluted 1:10 to 5 µM in cold 1x T4 ligase buffer. Approximately 8 µg of precipitated NotI digested gDNA was brought to 100 µL in autoclaved ddH_2O. To this 100 µL of DNA, 15 µL of 10x T4 ligase buffer, 3 µL of T4 ligase, 4.5 µL of annealed adaptors (0.15M), and 27.5 µL of autoclaved ddH_2O were added. The ligation reaction was incubated in a 16°C water bath overnight. The next day, the large gDNA protocol as described above was repeated.

Illumina Sample Preparation – MmeI Digestion – 5 µg of precipitated DNA was diluted to 50 µL in autoclaved ddH_2O. The M12 oligos (Tm332/333) were annealed as described for the adaptors in the previous step. To the 50 µL of DNA, 125.2 µL of ddH_2O, 20 µL of 10x CutSmart Buffer 0.8 µL of SAM (32mM), 2 µL of M12 dsDNA (50 µM), and 2 µL MmeI were added and the reaction mixture was incubated for 2 hours at 37°C. The DNA was then stored at -80°C.

Illumina Sample Preparation – Binding to Streptavidin Dynabeads® - 32 µL of streptavidin-coated Dynabeads® per sample were placed in a magnetic particle collector (MPC) for 1-2 minutes. The bead storage supernatant was removed and replaced with 1 mL 1x B and W Buffer (2x: 2M NaCl, 10mM Tris, 1mM EDTA, pH 7.5). This wash was repeated twice more before the beads were resuspended in 32 µL of 1x B and W Buffer per sample. To the MmeI digested DNA, 200 µL of 2x B and W Buffer was added along with 32 µL of beads per sample. This mixture was incubated with inversion for 15 minutes before being placed back in the MPC to remove the

supernatant and washed with 1 mL LoTE Buffer (3mM Tris, 0.2mM EDTA, pH 7.5) + 0.05% Tween. The beads were washed twice more with 1mL LoTE without Tween. Finally, the beads were resuspended in 100 µL LoTE and transferred to PCR tubes. The beads were stored at -4°C.

Illumina Sample Preparation – Annealing Barcode Adaptors – Illumina multiplex barcode adaptors were annealed as described above. One barcode is used per sample. The annealed barcode adaptors were diluted 1:10 to 5 µM in cold 1x T4 ligase buffer. A ligation reaction buffer mixture was prepared with 16.4 µL of ddH_2O and 2 µL of 10x T4 ligase buffer per sample. The Dynabeads® from the previous step were collected in the MPC, the buffer was removed, and the ligation reaction mixture was added to resuspend the beads. To this, 0.6 µL of the annealed 5 µM barcode adaptors was added along with 1 µL of T4 ligase. The reaction was mixed by pipetting and incubated for 7 hours at 16°C with gently mixing every hour. Afterward, 150 µL of LoTE + 0.05% Tween was added, the beads were collected, and the buffer was removed. This was repeated with 150 µL of LoTE without Tween. Finally, the beads were suspended in another 150 µL of LoTE without Tween and the beads were stored at 4°C overnight.

Illumina Sample Preparation – Final Amplification – The buffer was removed from the beads with the MPC and replaced with the following PCR reaction mixture: 5 µL 10x KOD buffer, 3 µL $MgSO_4$ (25mM), 5 µL dNTPs (2mM), 35.5 µL ddH_2O, 0.5 µL primer Tm1154 (Illumina P7), 0.5 µL primer Tm1178 (Illumina P5), and 0.5 µL KOD Hot Start Polymerase. The reaction was run with the following thermocycler conditions: 95°C for 2 minutes, 95°C for 20 seconds (cycle start), 60°C for 20 seconds, 72°C for 20 seconds (cycle end, 16 cycles), 12°C hold. The reaction was run on a 2% agarose gel in TAE with 1x light-colored loading dye for approximately 20 minutes. The

gel bands were excised and extracted. The DNA concentration of the final sequencing preparations was measured and the samples were stored at -80°C until submission.

Bioinformatics – The Geneious software by Biomatters Limited was used to trim and align the Tn-seq data (39). BLAST was used for all local alignments of significant genes.

Chapter 6

Conclusion and Future Directions

The system described herein is a powerful tool for examining the global regulators and alleles responsible for divergent phenotypes between bacterial strains. We applied this methodology to examine differences in expression of LPS modifications between *E. coli* K-12 and *E. coli* BL21(DE3), referred to as *E. coli* B. These modifications play a role in chronic immune responses to the products of the human microbiome. I believe our research in understanding this divergent genetic regulation of the microbiome could pave the way toward diagnosing and developing therapies for metabolic endotoxemia and its associated complications. In the era of increasingly personalized medicine, this knowledge could be applied to drug development from a pharmacogenomics viewpoint in order to optimize therapeutic agents for a patient's specific microbiome profile.

Simply by manually choosing the two most significant peaks in the sequencing data, we have identified several putative polymorphisms that could be responsible for the divergent LPS phenotype between *E. coli* K-12 and *E. coli* B. This has allowed us to develop two theories for mechanism of divergence (Figure 17). Firstly, we hypothesize that a hypoactivate allele in *feoB* contributes to increased expression of PEtN modifications. The cell takes in ferric iron (Fe^{3+}) from its surroundings via siderophores that transport it across the outer membrane (40). Once inside the periplasm, this iron can undergo reduction to ferrous iron (Fe^{2+}) which can be shuttled into the cytoplasm by the inner membrane *feoB* transporter. There are many iron transport operons within *E. coli*, but *feoB* is believed to be the most widespread (41). If any of the SNPs that we have

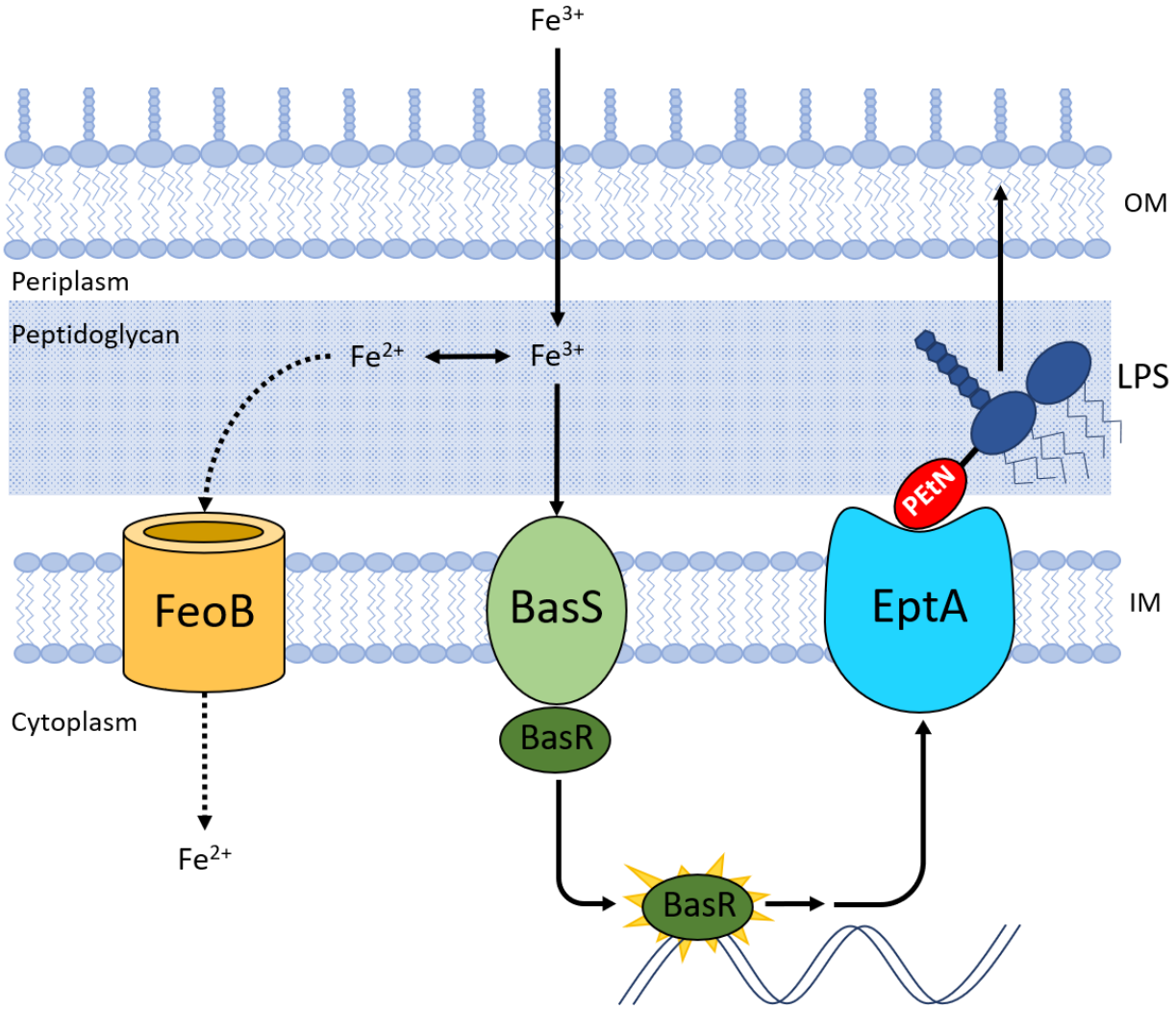


Figure 17: Hypotheses for phenotypic divergence

This pathway describes how the SNPs found in significant genes from the Tn-seq analysis could result in upregulation of PEtN modification in *E. coli* B. If the B allele of the inner membrane transporter *feoB* results in hypoactivity, then the periplasm will have an increased concentration of iron to activate *basSR*. Likewise, if the B alleles of *basS* or *basR* are hyperactive, then *eptA* will be overly expressed through transcriptional activation via *basR*.

identified in *feoB* result in hypoactivity within *E. coli* B, then there would be an increase in periplasmic iron concentration. The oxidations state of iron changes with oxygen availability and pH. Ferric reductases can convert the iron to Fe^{2+} while aerobic conditions will shift the equilibrium toward Fe^{3+} (42). An accumulation of Fe^{3+} in the periplasm due to a reduction of *feoB*

activity will overstimulate *basS* which senses periplasmic Fe^{3+} (43). This will upregulate the downstream effector *basR* to engage expression of *eptA* (44).

The second mechanism that we uncovered is that there is likely an uncharacterized activating allele in *E. coli* B that is likely located within *basSR*. *BasS* directly senses periplasmic iron through binding with an ExxE amino acid motif (12, 43). Binding induces *basS* to phosphorylate its two-component partner, *basR*, which dissociates from *basS* to act as a DNA-binding transcription factor (12, 44, 45). The promoter for *eptA* is a known target. Allelic differences in either the iron binding domain or the intracellular kinase domain could result in hyperactivation in *E. coli* B. As stated above, some mutations in *basR* that result in hyperactivation have already been characterized (3, 30, 31) and suggest that natural allelic differences could result in the same phenotype. How the specific activating SNPs we found—which differ from any so far characterized alleles—activate *basSR* in *E. coli* B is unknown.

The next goal of this project is to resume progress on the *arnT* reporter, MYAS590, and complete the transduction of the transposon library. Following NGS, the same analysis will be performed to find genetic regulators of the *arnT* gene. We suspect that there is an issue with the compatibility between the intrinsic strength of this reporter compared to *eptA* and the allele of resolvase used in the reporter. The initial assay to determine the allele of resolvase to use was only performed under the *eptA* promoter. Re-cloning the *arnT* reporter with a different resolvase allele could better facilitate the reporter's analysis.

We would also like to validate the conclusions from our NGS data. The connection between *feoB* and LPS modification could represent a novel link between the iron metabolism of gut microbiota and chronic inflammation. Initially, we would like to test our reporter strain under

varying concentrations of iron to assay induction. Then, we would like to generate *feoB* knockouts and overexpression models to test for LPS modification induction.

While the system was sophisticated enough to detect phenotypic differences resulting from whole-gene transfer between genetically-distinct strains, it could be enhanced to detect subtler differences in gene expression. The transposon-sequencing technique employed by Santiago, et al. makes use of multiplex sequencing and several distinct transposons that contain elements such as forward and reverse promoters and transcriptional terminators (46). These elements detect phenotypic changes resulting from gradations of upregulation and downregulation, not just single allele differences. The model organisms under examination could also be modified to facilitate the use of our system. The strains of *E. coli* in use in this study are phylogenetically similar (47). This allows the simple transfer of DNA between organisms. However, due to the bacterial restriction system, a form of prokaryotic immunity against foreign DNA, more divergent strains would not so easily incorporate DNA from a donor library. Elimination of the restriction system would be necessary to apply this system to phylogenetically dissimilar strains.

Our reporter model could theoretically be applied to analyze natural allelic differences in any gene in *E coli*. Thus, it represents a powerful tool for the study of microbial genomes. However, many questions remain about the impact of LPS on human patients of chronic inflammatory disease.

Appendix

Supplementary Information

Table 4: List of Strains

Strain Number	Strain Origin	Genotype
MYAS365	*E. coli* MC1000	Strain K10447 from Dr. David Friedman - *galK::PcatR-resC-tet-resC-catR*
TXM338	*E. coli* BW25113	E. coli BW25113 wild-type K-12 lineage
MYAS382	*E. coli* BW25113	SIVET - *galK::resC-tet-resC*
MYAS389	Stellar *E. coli*	pASCwt (pUC19-*eptA-tnpRwt-lacZ*)
MYAS390	Stellar *E. coli*	pASC168 (pUC19-*eptA-tnpR168-lacZ*)
MYAS391	Stellar *E. coli*	pASC135 (pUC19-*eptA-tnpR135-lacZ*)
MYAS392	Stellar *E. coli*	pASHwt (pUC19-*arnT-tnpRwt-lacZ*)
MYAS393	Stellar *E. coli*	pASH168 (pUC19-*arnT-tnpR168-lacZ*)
MYAS394	Stellar *E. coli*	pASH135 (pUC19-*arnT-tnpR135-lacZ*)
MYAS436	*E. coli* BW25113	*galK*:: SIVET, pSEVA::*eptA-tnpRwt*
MYAS437	*E. coli* BW25113	*galK*:: SIVET, pSEVA::*eptA-tnpR135*
MYAS439	*E. coli* BW25113	*galK*:: SIVET, pSEVA::*eptA-tnpR168*
MYAS491	*E. coli* Pir-1	pTNS2
MYAS492	*E. coli* Pir-1	pUC18R6K mini-Tn7T-Gm
MYAS510	*E. coli* BW25113	*galK*:: SIVET, *glmS::eptA-tnpR135-lacZ-GM*
MYAS590	*E. coli* BW25113	*galK*:: SIVET, *glmS::arnT-tnpR135-lacZ-GM*
TXM319	*E. coli* BL21DE3	wild-type E. coli B lineage
MYAS591	*E. coli* B (TM 319)	P1-vir from MYAS 510: *eptA-tnpR135-lacZ*, GM
MYAS604	*E. coli* B (TM 319)	P1-vir from MYAS 590: *arnT-tnpR135-lacZ*, GM
MYAS654	Stellar *E. coli*	pKD46 + C9HMAR In-Fusion+transformed
MYAS664	*E. coli* Pir-116	*Tn:kanR-cos-oriT-oriR6K* plasmid
MYAS667	*E. coli* BL21DE3	pKD46 + C9HMAR

Table 5: List of Plasmids

Plasmid	Genotype	Origin
pKD46	*araC-bla-γ-β-exo*	Datsenko and Warner (48)
pGOA1193	*oriR6K mobRP4 tnpR lacZ Apr*	Osorio et al. (16)
pGOA1194	*oriR6K mobRP4 tnpR*^{mut168} *lacZ ApR*	Osorio et al. (16)
pGOA1195	*oriR6K mobRP4 tnpR*^{mut135} *lacZ ApR*	Osorio et al. (16)
pUC19	*ApR*	Lab
pASCwt	*eptA-tnpRwt-lacZ*	This work
pASC168	*eptA-tnpR168-lacZ*	This work
pASC135	*eptA-tnpR135-lacZ*	This work
pASHwt	*arnT-tnpRwt-lacZ*	This work
pASH168	*arnT-tnpR168-lacZ*	This work
pASH135	*arnT-tnpR135-lacZ*	This work
pSEVAwt	*eptA-tnpRwt*	This work
pSEVA135	*eptA-tnpR135*	This work
pSEVA168	*eptA-tnpR168*	This work
pTNS2	*oriR6K bla tnsABCD*	Choi et al. (19, 20)
pUC18R6K-mini-Tn7T-Gm	*bla* GmR Tn7R-Tn7L	Choi et al. (19, 20)
pWV01	*kanR* rbs C9 HMAR	Lab
pKD4	*kanR-oriRγ-bla*	Datsenko and Warner (48)
pWEB	*NeoR-AmpR-cos*	Epicentre® (49)
pTM223	Tn cassette	Lab (46)
pKD3	*catR-oriRγ-bla*	Datsenko and Warner (48)
pEX18	*oriT-sacB-AmpR*	Lab (50)

Table 6: List of Primers

Primer Name	Sequence
AS553-p1 SIVET into *galK*	GTTAGCCGTCGTGCCTTACTGG
AS554- p2 SIVET into *galK*	CGGCCTGATCCTGATAGCATTC
AS561 *tnpR* FOR BglII	CTAGAATGTCGGCTTTTTCTCG
AS562 *tnpR lacZ* REV HindIII	CTTGCAAAGCTTGGAAAGCCTGGTCGGCTG
Tm466-*eptA* for EcorI	GACGGCCAGTGAATTCACCCGTATCCCTTAG
Tm467-*eptA* rev BamHI	CTAACAGGTTGGATCCCACGGTGTTTCCATC
Tm450-*arnT* for EcoRI	GACGGCCAGTGAATTCACTCCACCTATAGAC
Tm465-*arnT* pMYAS rev BamHI	CTAACAGGTTGGATCCTGCTTTTCCTTCCGC
Tm727-pUC19 *eptA tnp lacZ* XhoI for	GTTAGTCTCGAGGAATTCACCCGTATCC
Tm728- pUC19 *arnT tnp lacZ* XhoI for	GTTAGTCTCGAGGAATTCACTCCACCTATAG
Tm729- pUC19 promoter *tnp lacZ* BamHI rev	GTTAGTGGATCCGAAAGCCTGGTCGGCTG
TM903- C9HMAR for EcoRI	GTTTTTTTGGGAATTCGGATGAATTGTTTTAGTAC
TM904- C9HMAR rev NcoI	CTGTCCATACCCATGGATAAACAAAATGATCAAGC
TM909-pKD46 backbone rev EcoRI	AATTCCCAAAAAAACGGGTATGG
TM910-pKD46 backbone for NcoI	CATGGGTATGGACAGTTTTCC
Tm214-biotin NotI tag	CGTTAGTAACCTTGCGATGTCGATTCACGTTG
Tm215-NotI tag	GGCCCAACGTGAATCGACATCGCAAGGTTACTAACG
Tm332-M12_top	CTGTCCGTTCCGACTACCCTCCCGAC
Tm333-M12_bottom	GTCGGGAGGGTAGTCGGAACGGACAG
TM916-*kanR* for BspHI	GAATTAGCTTTCATGCAGTGGGCTTACATGGC
TM917-*kanR* rev BsrGI	TTATCCTTATTGTACATCGAAATCTCGTGATGG
Tm294- *Cos* for	TTGATCTTTTCTACGGATCTGCCTCGCTGG
Tm295- *Cos* rev	GTCTGGATCCTTTGCCGATAGTATGCAATTG
Tm290- Tn Gm cassette for	CCGCATTAAAGCTTAACAGCGCATGCCAAG
Tm291- Tn Gm cassette rev	GCAAAGGATCCAGACTACTG
TM931- *pir ori* for	CGTAGAAAAGATCAACCATGGCTAATTCCC
TM932- *pir ori* rev	GAGGCTGGCCGGCTACAAGATCCGCAGTTC
TM933-*oriT* for	TAGCCGGCCAGCCTCGCAGAG
TM934-*oriT* rev	TAAGCTTTAATGCGGGCATAACCCTGCTTC

BIBLIOGRAPHY

1. Neves AL, Coelho J, Couto L, Leite-Moreira A, Roncon-Albuquerque R. 2013. Metabolic endotoxemia: A molecular link between obesity and cardiovascular risk. J Mol Endocrinol 51.
2. Greenblum S, Turnbaugh PJ, Borenstein E. 2012. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and in fl ammatory bowel disease. Proc Natl Acad Sci 109:594–599.
3. Nummila K, Kilpeläinen Ii, Zähringer U, Vaara M, Helander IiM. 1995. Lipopolysaccharides of polymyxin B-resistant mutants of Escherichia coii are extensively substituted by 2-aminoethyl pyrophosphate and contain aminoarabinose in lipid A. Mol Microbiol 16:271–278.
4. Moreira APB, Texeira TFS, Ferreira AB, Do Carmo Gouveia Peluzio M, De Cássia Gonçalves Alfenas R. 2012. Influence of a high-fat diet on gut microbiota, intestinal permeability and metabolic endotoxaemia. Br J Nutr 108:801–809.
5. Mogensen TH. 2009. Pathogen recognition and inflammatory signaling in innate immune defenses. Clin Microbiol Rev 22:240–273.
6. Murphy EA, Velazquez KT, Herbert KM. 2015. Influence of High-Fat-Diet on Gut Microbiota: A Driving Force for Chronic Disease Risk. Curr Opin Clin Nutr Metab Care 18:515–520.
7. Maeshima N, Fernandez RC. 2013. Recognition of lipid A variants by the TLR4-MD-2 receptor complex. Front Cell Infect Microbiol 3:1–13.

8. Lu YC, Yeh WC, Ohashi PS. 2008. LPS/TLR4 signal transduction pathway. Cytokine 42:145–151.

9. Cani PD, Possemiers S, Van De Wiele T, Guiot Y, Everard A, Rottier O, Geurts L, Naslain D, Neyrinck A, Lambert DM, Muccioli GG, Delzenne NM. 2009. Changes in gut microbiota control inflammation in obese mice through a mechanism involving GLP-2-driven improvement of gut permeability. Gut 58:1091–1103.

10. Boutagy NE, Mcmillan RP, Frisard MI, Hulver MW, Translational F, Tech V, Core MP, Tech V. 2016. Metabolic endotoxemia with obesity: is it real and is it relevant? Biochimie 11–20.

11. Rubin EJ, Herrera CM, Crofts AA, Trent MS. 2015. PmrD is required for modifications to escherichia coli endotoxin that promote antimicrobial resistance. Antimicrob Agents Chemother 59:2051–2061.

12. Chen HD, Groisman EA. 2013. The Biology of the PmrA/PmrB Two-Component System: The Major Regulator of Lipopolysaccharide Modifications. Annu Rev Microbiol 67:83–112.

13. Herrera CM, Hankins J V., Trent MS. 2010. Activation of PmrA inhibits LpxT-dependent phosphorylation of lipid A promoting resistance to antimicrobial peptides. Mol Microbiol 76:1444–1460.

14. Lee H, Hsu F, Turk J, Groisman EA. 2004. The PmrA-Regulated pmrC Gene Mediates Phosphoethanolamine Modification of Lipid A and Polymyxin Resistance in Salmonella enterica 186:4124–4133.

15. Winfield MD, Groisman EA. 2004. Phenotypic differences between Salmonella and Escherichia coli resulting from the disparate regulation of homologous genes. Proc Natl

Acad Sci 101:17162–17167.

16. Osorio CG, Crawford JA, Michalski J, Kaper JB, Camilli A. 2005. In Vivo Expression Technology for Large-Scale Screening for Vibrio cholerae Genes Induced during Infection of the Mouse Small Intestine Second-Generation Recombination-Based In Vivo Expression Technology for Large-Scale Screening for Vibrio cholerae Genes. Infect Immun 73:972–980.
17. Livny J, Friedman DI. 2004. Characterizing spontaneous induction of Stx encoding phages using a selectable reporter system. Mol Microbiol 51:1691–1704.
18. Merighi M, Ellermeier CD, Slauch JM, Gunn JS. 2005. Resolvase-In Vivo Expression Technology Analysis of the Salmonella enterica Serovar Typhimurium PhoP and PmrA Regulons in BALB / c Mice †. Society 187:7407–7416.
19. Choi KH, Gaynor JB, White KG, Lopez C, Bosio CM, Karkhoff-Schweizer RAR, Schweizer HP. 2005. A Tn7-based broad-range bacterial cloning and expression system. Nat Methods 2:443–448.
20. Choi KH, Schweizer HP. 2006. mini-Tn7 insertion in bacteria with single attTn7 sites: Example Pseudomonas aeruginosa. Nat Protoc 1:153–161.
21. New England BioLabs Inc. 2018. PCR Using Q5 ® High-Fidelity DNA Polymerase (M0491). www.neb.com.
22. New England BioLabs Inc. 2018. Ligation Protocol with T4 DNA Ligase (M0202). www.neb.com.
23. Zhang X, Bremer H. 1995. Control of the Escherichia coli rrnB P1 Promoter Strength by ppGpp*. J Biol Chem 270:11181–11189.
24. Miller JH. 1972. Experiments in molecular genetics. Cold Spring Harb Lab Press Cold

Spring Harb NY 433:352–355.

25. Meredith TC, Wang H, Beaulieu P, Gründling A, Roemer T. 2012. Harnessing the power of transposon mutagenesis for antibacterial target identification and evaluation. Mob Genet Elements 2:171–178.

26. Freddolino PL, Goodarzi H, Tavazoiea S. 2014. Revealing the genetic basis of natural bacterial phenotypic divergence. J Bacteriol 196:825–839.

27. Takara Bio USA. 2016. In-Fusion® HD Cloning Kit User Manual.

28. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IMJ, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar S V., Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DMD, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Catenazzi MCE, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo K V., Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov D V., Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall

AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59.

29. Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A Greedy Algorithm for Aligning DNA Sequences. J Comput Biol 7:203–214.

30. Froelich JM, Tran K, Wall D. 2006. A pmrA Constitutive Mutant Sensitizes Escherichia coli to Deoxycholic Acid A pmrA Constitutive Mutant Sensitizes Escherichia coli to Deoxycholic Acid 188:1180–1183.

31. Helander IM, Kilpeläinen I, Vaara M. 1994. Increased substitution of phosphate groups in lipopolysaccharides and lipid A of the polymyxin-resistant pmrA mutants of Salmonella typhimurium: a 31P-NMR study. Mol Microbiol 11:481–487.

32. Kammler M, Schön C, Hantke K. 1993. Characterization of the ferrous iron uptake system

of Escherichia coli. J Bacteriol 175:6212–6219.

33. Kingston AW, Ponkratz C, Raleigh EA. 2017. Rpn (YhgA-Like) Proteins of Escherichia coli K-12 and Their Contribution to RecA-Independent Horizontal Transfer Anthony. J Bacteriol 199:1–17.

34. Lin S, Cronan JE. 2011. Closing in on complete pathways of biotin biosynthesis. Mol Biosyst 7:1811.

35. Porco A, Alonso G, Isturiz T. 1998. The gluconate high affinity transport of gntI in Escherichia coli involves a multicomponent complex system. J Basic Microbiol 38:395–404.

36. Angelini S, Gerez C, Choudens SO De, Sanakis Y, Fontecave M, Barras F, Py B. 2008. NfuA, a new factor required for maturing Fe/S proteins in Escherichia coli under oxidative stress and iron starvation conditions. J Biol Chem 283:14084–14091.

37. Porco A, Peekhaus N, Bausch C, Tong S, Isturiz T, Conway T. 1997. Molecular genetic characterization of the [i]Escherichia coli[/i] [i]gntT[/i] gene of GntI, the main system for gluconate metabolism. J Bacteriol 179:1584–1590.

38. Ilgü H, Jeckelmann J-M, Gapsys V, Ucurum Z, de Groot BL, Fotiadis D. 2016. Insights into the molecular basis for substrate binding and specificity of the wild-type L-arginine/agmatine antiporter AdiC. Proc Natl Acad Sci 113:10358–10363.

39. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

40. Chu BC, Garcia-Herrero A, Johanson TH, Krewulak KD, Lau CK, Peacock RS,

Slavinskaya Z, Vogel HJ. 2010. Siderophore uptake in bacteria and the battle for iron with the host; a bird's eye view. BioMetals 23:601–611.

41. Lau CKY, Ishida H, Liu Z, Vogel HJ. 2013. Solution Structure of Escherichia coli FeoA and Its Potential Role in Bacterial Ferrous Iron Transport. J Bacteriol 195:46–55.
42. Rong C, Huang Y, Zhang W, Jiang W, Li Y, Li J. 2008. Ferrous iron transport protein B gene (feoB1) plays an accessory role in magnetosome formation in Magnetospirillum gryphiswaldense strain MSR-1. Res Microbiol 159:530–536.
43. Wösten MMSM, Kox LFF, Chamnongpol S, Soncini FC, Groisman EA. 2000. A signal transduction system that responds to extracellular iron. Cell 103:113–125.
44. Ogasawara H, Shinohara S, Yamamoto K, Ishihama A. 2012. Novel regulation targets of the metal-response BasS-BasR two-component system of Escherichia coli. Microbiol (United Kingdom) 158:1482–1492.
45. Shin D, Groisman EA. 2005. Signal-dependent binding of the response regulators PhoP and PmrA to their target promoters in vivo. J Biol Chem 280:4089–4094.
46. Santiago M, Matano LM, Moussa SH, Gilmore MS, Walker S, Meredith TC. 2015. A new platform for ultra-high density Staphylococcus aureus transposon libraries. BMC Genomics 16:1–18.
47. Sims GE, Kim S-H. 2011. Whole-genome phylogeny of Escherichia coli/Shigella group by feature frequency profiles (FFPs). Proc Natl Acad Sci 108:8329–8334.
48. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–5.
49. Epicentre. pWEB^{TM} Cosmid Cloning Kit.
50. Yuan Y, Sachdeva M, Leeds JA, Meredith TC. 2012. Fatty acid biosynthesis in

Pseudomonas aeruginosa is initiated by the faby class of β-ketoacyl acyl carrier protein synthases. J Bacteriol 194:5171–5184.

ACADEMIC VITA

Alexander B. Smith

alexsmith5487@gmail.com

Education

B.S. in Biochemistry and Molecular Biology **May 2018**

Pennsylvania State University, Schreyer Honors College, State College, PA

B.S. Honors Thesis *(*Advisor: Dr. Timothy Meredith)

Genome-Wide Screen for Induction of Phosphoethanolamine and Aminoarabinose Modifications in E. coli LPS

Elucidate the genetic mechanism underlying differential *E. coli* lipopolysaccharide modifications in the human gut microbiome as it relates to chronic inflammatory diseases.

Education Abroad **January-May 2017**

University College Dublin, Dublin, Ireland

Research Experience

Summer Undergraduate Research in Pharmacology **June-August 2017**

Laboratory of Dr. Jen-Tsan Ashley Chi, Duke University

- Generate knockout cell lines using CRISPR/Cas9
- Clone and express tagged-proteins for study of nutrient-sensing in cancer environments
- Investigate role of glycosylation in tumor suppressor regulatory pathways

Undergraduate Researcher **January 2015-Present**

Laboratory of Dr. Timothy Meredith, Pennsylvania State University

- Design, clone unique DNA construct for reporting select lipopolysaccharide modifications
- Adapt classic DNA cloning techniques to support complex recombinant DNA projects
- Built transposon library for transposon sequencing of constructed reporter strain via NGS

Undergraduate Researcher **January-May 2017**

Laboratory of Dr. Patrick Caffrey, University College Dublin

- Clone, express, and purify a polyene glycosyltransferase enzyme in *E. coli*
- Test the solubility of this glycosyltransferase to aid in the bacterial production of less-toxic antifungal compounds.

Honors, Awards, and Grants

Barry M. Goldwater Scholarship	2017
ASM Undergraduate Research Fellowship	2017
Herko Family Scholarship in Biochemistry and Molecular Biology	2017
Evan Pugh Scholar Senior Award	2017
Pennsylvania Space Grant Consortium Undergraduate Scholarship	2017
Venezie Scholarship in Science for Honors Education	2016
Erickson Discovery Grant	2016

Schreyer International Travel Grant	2016
Rosenthal Family Science Scholarship	2016
President Sparks Award	2016
President's Freshmen Award	2015
Academic Excellence Scholarship	2014-2017

Presentations

Smith, A., Yi, M., Komazin, G., Meredith, T. (2017, October). *Genome-Wide Screen for Induction of Phosphoethanolamine and Aminoarabinose Modifications in E. coli LPS.* Allegheny Branch of the American Society of Microbiology, Juniata College, Huntingdon, PA.

- Third Place in Poster Competition

Smith, A., Caffrey, P. (2017, October). *Overproduction of Polyene Macrolide Glycosyltransferases in Escherichia Coli.* Eberly College of Science Benefactor's Dinner, University Park, PA.

Smith, A. (2017, August). *Characterizing O-GlcNAcylation in the KLHL Family.* Summer Undergraduate Research in Pharmacology and Cancer Biology, Durham, NC.

Smith, A., Caffrey, P. (2017, May). *Overproduction of Polyene Macrolide Glycosyltransferases in Escherichia Coli.* SCI30010 Final Poster Presentation, University College Dublin, Dublin, Ireland.

Smith, A., Yi, M., Komazin, G., Meredith, T. (2016, September). *Genome-Wide Screen for Induction of Phosphoethanolamine and Aminoarabinose Modifications in E. coli LPS.* Eberly College of Science Poster Exhibition, University Park, PA.

Smith, A., Yi, M., Komazin, G., Meredith, T. (2016, April). *Genome-Wide Screen for Induction of Phosphoethanolamine Modifications in E. coli LPS.* Undergraduate Research Exhibition, University Park, PA.

Leadership, Involvement, and Volunteering

Schreyer Honors College Student Council **September 2015-Present**

Recruitment Chair, Active Member

- Direct receptions and tours of up to 250 prospective students and their parents for the honors college
- Introduce the Schreyer Honors College to a record number of applicants
- Advance the mission of the honors college by participating in recruitment, service, and leadership programming

Discovery Space of Central Pennsylvania **October 2017 – Present**

Floor Exhibit Volunteer

- Maintain the exhibits and ensure that all stations are tidy and in good working condition
- Facilitate interactions between parents, children, and the exhibits to promote science education in the community

Science LionPride **September 2014 – Present**
Active Member

- Tour prospective students as an ambassador of the Eberly College of Science
- Attend alumni networking functions to foster relationships between students and alumni

Schreyer Honors Orientation **March 2015-August 2016**
Team Leader, Mentor

- Managed a group of eleven mentors who guided roughly fifty mentees through orientation
- Coordinated orientation programming for incoming Schreyer Scholars

Teaching Experience

Teaching Assistant, Introductory Microbiology **Spring 2018**
Department of Biochemistry and Molecular Biology, Pennsylvania State University

- Assisted in the development of in-class assignments to supplement the lectures
- Facilitated in-class discussion and answered student questions

Tutor, Fundamentals of Organic Chemistry I **Fall 2015**
Department of Chemistry, Pennsylvania State University

- Facilitated both large and small group review discussions on organic chemistry
- Communicated with students to develop personalized tutoring strategies

Professional Affiliations

American Society of Microbiology **February 2017-Present**
Member