Characterizing the distribution of epistatic effects using a
sequence-based assay in *Escherichia coli*

by

Melissa Kohlman

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements for the degree of

Master of Science

in

Biology

Carleton University
Ottawa, Ontario

© 2014, Melissa Kohlman

Abstract

Epistasis, the modification of the fitness effects of a mutation by other mutations, has important consequences for the trajectory of adaptive evolution and for the shape of the “adaptive landscape”. Nonetheless, the extent of epistasis, and the shape of the distribution of epistatic effects (DoEE), are poorly characterized. We developed a high-throughput method for describing the DoEE. A set of 4123 *E. coli* ORFs (the ASKA collection) was transformed into a strain bearing a common antibiotic resistance mutation, and into a wild-type strain. Upon induction of expression, the fitness of each ORF was determined using a sequence-based assay, providing a rapid measure of background-dependent fitness effects for over 4000 constructs. 216 epistatic interactions were detected. Synthetic dosage lethal/sick interactions from the over-expression of *yiaN* and *nfi* were validated using 24 hour growth curves, and may indicate that these genes are potential targets for drug design.

Acknowledgements

I would first like to thank Dr. Alex Wong for taking me under his wing for the last 3 years in all things having to do with science and life, first during my Honour's project at the University of Ottawa, and then welcoming me to Carleton University to do a Master's degree. His brilliance, wit, patience, endless supply of chocolate and making sure I never gave up to become a hippy in Tofino, has greatly expanded my scientific skills and mind, while keeping my ego somewhat in check. I would like to thank Dr. Myron Smith for agreeing to be my co-supervisor before he even met me, and by recommending I extract plasmids in batches, saving me from a summer of plasmid hell! Thank you Dr. Cathy Carrillo and Dr. Alex Poulain for advising me through very warm and helpful committee meetings. I would like to thank my lab for always being ready to help out (Liam O'Sullivan!), constantly cheering me on (Prabh, Ahlam, Kamya, Tom, Jess, Ian, Trevor, Ryan, Abi), refraining from stealing my stuff, dealing with my grumpy sarcasm, and listening to my really loud music. Finally I would like to thank my family: Doug for reminding me I am his smart big sister and Dad, you never ever stopped believing in me, despite only knowing I am "trying to save the world" by conquering antibiotic resistance, while having no clue how.

Mom, this one's for you.

Table of Contents

Abstract.....	ii
Acknowledgements	iii
Table of Contents	v
List of Tables.....	viii
List of Figures.....	ix
Introduction	1
<i>Epistasis.....</i>	<i>4</i>
<i>What is Epistasis?</i>	<i>4</i>
<i>Influences of Epistasis</i>	<i>7</i>
<i>High-Throughput Screens</i>	<i>9</i>
<i>Synthetic Lethal/Synthetic Dosage Lethal Interactions.....</i>	<i>9</i>
<i>How Synthetic Lethal/Synthetic Dosage Lethal Interactions were previously identified.....</i>	<i>10</i>
<i>Distribution of Epistatic Interactions determined from previous screens</i>	<i>13</i>
<i>Using Next Generation Sequencing to characterize Synthetic Dosage Lethal Interactions..</i>	<i>14</i>
<i>Evolution and Epistasis</i>	<i>15</i>
<i>Antibiotic Resistance</i>	<i>17</i>
<i>Widespread resistance/Impact on Community</i>	<i>18</i>
<i>Resistance in E. coli</i>	<i>20</i>
<i>Using epistasis- finding new drug targets through a high-throughput screen.....</i>	<i>21</i>
Materials and Methods	22

<i>Construction of the quinolone resistant mutant genotype.....</i>	<i>22</i>
<i>Construction of the gyrA S83L allele</i>	<i>22</i>
<i>Gene gorging.....</i>	<i>23</i>
<i>Plasmid Library Replication</i>	<i>26</i>
<i>Plasmid Library Extraction.....</i>	<i>27</i>
<i>High-throughput screen for genetic interactions</i>	<i>28</i>
<i>Pre-Sequencing Preparation.....</i>	<i>29</i>
<i>Sequencing</i>	<i>30</i>
<i>Bioinformatics and Data Analysis.....</i>	<i>30</i>
<i>Bioinformatics workflow</i>	<i>30</i>
<i>Data and Statistical analysis</i>	<i>31</i>
<i>Epistatic interaction validation.....</i>	<i>34</i>
Results.....	35
<i>High-throughput screen for genetic interactions</i>	<i>35</i>
<i>Sequencing and bioinformatics workflow</i>	<i>35</i>
<i>Sample of sequence counts using HTSeq</i>	<i>38</i>
<i>Sample of data and statistical analysis</i>	<i>40</i>
<i>Evaluation of consistency between trials</i>	<i>43</i>
<i>Potential epistatic interactions.....</i>	<i>44</i>
<i>Gene Validation</i>	<i>46</i>
<i>Synthetic dosage sick/lethal gene interactions for the S83L gyrA mutant.....</i>	<i>46</i>
<i>Beneficial gene interactions for the S83L gyrA mutant.....</i>	<i>49</i>
<i>Validation Experiments</i>	<i>53</i>
Discussion.....	63

<i>High Throughput Method</i>	63
<i>Comparison to SGA</i>	63
<i>Challenges and Issues</i>	68
<i>Additional Uses- Evolution Experiments</i>	71
<i>Additional Uses- Antibiotic Drug Targets</i>	71
<i>Additional Uses- Cancer Drug Discovery</i>	72
<i>Additional Uses- Orphan Disease Treatment discovery</i>	73
<i>Discovery of Gene Candidates for Drug Targets and Beneficial Mutations</i>	74
<i>Antibiotic Resistance- A Growing Problem</i>	74
<i>Antibiotic Gene Candidates detected by Screen</i>	75
yiaN	75
nfi	77
<i>S83L-specific beneficial mutations</i>	78
puuB.....	79
ymdB.....	80
ynjH.....	80
hyfC.....	81
<i>Other potential drug targets</i>	82
<i>Comparison to current methods in antibiotic target discovery</i>	82
<i>Impact on Evolutionary Theory</i>	83
Conclusion.....	85
References	87
Glossary	94

List of Tables

Table 1: Sequence counts for the bioinformatic workflow.....	36
Table 2: The raw gene counts for each sample, determined by aligning reads to the reference genome.....	38
Table 3. Data analysis for the first 8 genes.....	39
Table 4: Synthetic dosage sick/lethal gene candidates on the mutant background	46
Table 5: Compensatory or beneficial gene candidates on the mutant background,	50
Table 6. Maximum growth rate <i>P</i> - values for tested genes.....	55
Table 7. Optical density <i>P</i> -values for tested genes at 6 hours.....	57
Table 8. Optical density <i>P</i> -values for tested genes at 24 hours.....	59
Table 9. Result of validation screen.....	61

List of Figures

Figure 1. A theoretical example of how epistasis can change fitness.....	2
Figure 2. The aligned sequences from the wild type and mutant <i>gyrA</i> gene.....	26
Figure 3. A visual depiction of the high-throughput screen.....	29
Figure 4. Bioinformatics workflow used to identify the genes, and the quantity present in the extracted plasmid.....	28
Figure 5. Selection coefficient correlations for wild type and mutant between the 3 trials.....	43
Figure 6. The distribution of the amount of genes showing epistatic interactions for the mutant strain at 6 hours.....	44
Figure 7. Fitness for all 4123 over-expression genes on the MG1655 wild type background versus the <i>gyrA</i> S83L mutant background after 6 hours.	45
Figure 8. Selection coefficients for 20 gene candidates that are synthetically dosage lethal/sick on the mutant S83L <i>gyrA</i> background at 6 and 24 hours.....	48
Figure 9. Selection coefficients for 20 gene candidates that are beneficial on the mutant S83L <i>gyrA</i> background at 6 and 24 hours.....	52
Figure 10. Optical density growth curves of liquid culture for 24 hours at 37 °C of the 1 confirmed genes <i>yiaN</i> and <i>nfi</i> ,.....	59
Figure 11. Optical density growth curves of liquid culture for 24 hours at 37 °C of the 4 confirmed genes <i>puuB</i> , <i>ynjH</i> , <i>hyfC</i> and <i>ymdB</i> ,	60

Introduction

Adaptation is the movement of a population towards a phenotype that best fits the environment it resides in, driven by beneficial mutations being selected for by natural selection (Orr, 2005). Focus has been placed on the mutations themselves, as little is known about the general characteristics of the mutations driving adaptation. Researchers have set out to determine the individual effects of mutations, asking for example whether adaptation is driven by weak frequent mutations, or rare larger mutations, or if the mutations arising in complex organisms differ from simpler organisms.

Instead of focusing individually on mutations that arise, we are interested in characterizing the extent that previously selected mutations have on future selected mutations. In other words, I aim to determine whether the starting genetic material shapes the subsequent evolutionary trajectory as a population evolves towards a higher fitness. If the fitness of any given mutation is independent of previous mutations, we expect to see little influence of starting genetic material on the final adapted population, other than providing the baseline fitness. However, if genetic or epistatic interactions are prevalent, then the initial genetic background may strongly influence which mutations are selected for, and hence the path of adaptation. Much is still unknown about the genetic basis of adaptation, with traditional mathematic models of evolution rarely addressing questions such as the extent of epistasis in shaping adaptation (Orr, 2005).

Adaptive walks, characterized by Maynard Smith, describe how mutations occur in a stepwise fashion. Beneficial mutations are selected for one after another, without passing through non-functional intermediates. Each step will contribute to improving a population's fitness until it reaches its ultimate fitness peak. The question is, how much does the previous step or mutation influence future steps when selecting for the next beneficial mutations? While it is known that fitness landscapes in specific cases such as HIV evolution are strongly influenced by genetic interactions, it is unknown how prevalent it is in protein evolution overall (Martinez, 2014).

Epistasis is defined as interactions between genes resulting in a phenotype that deviates from their combined independent effects (Figure 1). Evolutionarily, it is defined as a change in fitness that is not cumulative. This is best described using a bar graph, as seen in Figure 1. Mutations in genes 1 and 2, acting separately, may have a fitness increase of "2". With no epistasis, the mutations combined are expected to increase fitness to "4". If there is epistasis, then the cumulative effects of these mutations is no longer additive, and would therefore be greater or less than 4.

Figure 1. A theoretical example of how epistasis can change fitness.

Epistatic interactions are an important part of genetic pathways and therefore cellular function, in all organisms (de Visser, 2011). The extent to which these genetic interactions impact cellular function and shape evolutionary trajectories is still unknown. Better understanding of genetic interactions will help to shape answers about the influence of epistasis on evolution, genetic systems and even drug design. Developing high throughput screens to do so is necessary as there are limitless combinations of genetic interactions. High throughput screens that take advantage of next generation sequencing techniques may help to provide the speed and the processing power to look at thousands of gene interactions at once, while resulting in quantitative data that may more thoroughly describe epistatic effects better than previous screening methods. The application of such screens may provide us with new

ways to create drug targets against antibiotic resistant strains, create tailored chemotherapies against cancer, and predict how drug resistance in bacteria may arise.

Epistasis

What is Epistasis?

The term “epistasis” has been used to describe numerous related effects. Three separate interactions have fallen under the term epistasis: “the functional relationship between genes, genetic ordering of regulatory pathways and quantitative differences of allele-specific effects” (Phillips, 2009).

A functional relationship between genes describes the actual physical relationship that proteins have with each other. They may be in the same pathway or create protein complexes with one another (Phillips, 2009). Another term for this type of interaction is a protein-protein interaction, which is more descriptive. This type of interaction does not directly explain a genetic link.

Genetic ordering of regulatory pathways, or compositional epistasis, is what William Bateson was first describing when he coined the term epistasis. Bateson had observed phenotypes in a dihybrid cross that did not match the expected ratios. The result was one allele at a locus masking the phenotypic effects of another allele at a separate locus. He concluded that these genes must be interacting with each other, calling it epistasis. Epistasis translates to “standing upon”, as one gene masks another gene (which was termed the hypostatic locus) resulting in an unexpected phenotype

(Phillips, 1998). Such effects are seen in pea plant flower colours, the shape of chicken combs, and in fruit fly eye colour.

Fruit fly eye colour provides an excellent classic example of compositional epistasis in genetic pathways, with a phenotype that is easily interpreted. Fruit flies can have four eye colours: the combination of red and brown called brick red, red, brown and white. Two parallel pathways produce the typical brick red eye colour, which is composed of both the red and brown pigment (Mackenzie, 1999). Gene products from both the white and scarlet pathways make a tryptophan transporter, while gene products from both the white and brown pathways make a guanine transporter (Mackenzie, 1999). These transporters are essential in depositing the pigment in the eye. A mutation in the brown pathway creates a brown-pigmented eye, while a mutation in the scarlet pathway creates a scarlet-pigmented eye (Mackenzie, 1999). While the naming seems counterintuitive, the names reflect the outcome of the mutation, and not the actual pigment it is involved in producing. An epistatic interaction occurs when a mutation occurs in the white gene, upstream of both of the pathways. A knockout mutation creates a malfunctioning transporter protein and blocks the transportation and production of both pigments, producing a non-pigmented eye (Mackenzie, 1999). A white/brown double mutant or a white/scarlet double mutant still results in white eyes, as the white gene is upstream and “stands above” the brown and scarlet genes. This is a classic example of epistasis, as one mutation is blocking the effects of other alleles.

Compositional epistasis does not necessarily need to be between mutations at different loci. Intragenic epistasis occurs when mutations in the same gene have epistatic effects. The mutations at different nucleotide positions in a single gene, like mutations in separate genes, combine to have either positive or negative effects different than the sum of their individual effects (Martinez, 2014). Martinez and Parera showed the importance of intragenic epistasis in protein evolution by using site directed mutagenesis to change individual amino acids in a deleterious variant of a hepatitis C virus protease. The gene variant was deleterious on its own, but changing amino acids at other sites within the gene created lethal interactions or greatly improved fitness. This shows that single amino acid changes within a gene variant can have epistatic effects (Martinez, 2014).

Statistical epistasis, a term coined by Fisher, describes the effect of epistasis on all the gene combinations in populations rather than individuals (Phillips, 2009). The population mean fitness is measured, and within an individual changing the alleles at two loci. After determining the fitness effect changing each locus individually, if the combination of the two alleles results in a different fitness than each effect combined, epistasis has occurred. This type of epistasis takes into account variation in genetic backgrounds in a population, as there is no “standard background” (Phillips, 2009).

Compositional and statistical epistasis are similar. The difference between the two is the fitness standard that a change in phenotype or fitness from genetic interaction is measured by. Compositional epistasis measures epistasis as a change in phenotype due to the combination of two separate alleles different than expected on an

individual's background, whereas statistical epistasis measures the average affect of that change in comparison to the population's background. By looking at the average effect over an entire population, some epistatic interactions may be lost due to variation (Phillips, 2009).

Quantitative differences of allele-specific effects can be further broken down to two categories: unidimensional epistasis and multidimensional epistasis.

Unidimensional epistasis is determined by comparing the mean log fitness to the number of alleles affecting fitness and observing deviations (de Visser, 2011).

Unidimensional epistasis or directional epistasis can be either positive or negative. A deviation from the expected fitness is a result of epistasis. Positive unidimensional epistasis results from an increase in fitness more than expected from beneficial mutation. Negative unidimensional epistasis results from a decrease in fitness more than expected from a detrimental mutation. Multidimensional epistasis describes multiple interactions between mutations, resulting in mutations having different effects on different backgrounds (de Visser, 2011). Sign epistasis is an example of multidimensional epistasis, which will be discussed later.

Influences of Epistasis

Evolutionary changes are always built on existing pathways and genes in a step by step process, with gene interactions as a natural outcome (Phillips, 2009). Mutations that cause a loss of function of a specific gene may present more epistatic effects than a more subtle mutation. Often when studying epistatic effects, knockouts are used, and

just because a large change to the system results in an epistatic effect, the naturally occurring alleles at that locus may actually not be interacting epistatically (Phillips, 2009).

It has been found that mutations that arise have multiple effects on fitness and phenotype, meaning they are pleiotropic, and their effect on overall fitness greatly depends on the background that they arise on. As mutations arise, they may have both negative and positive effects, leading to selection of compensatory mutations to offset those negative effects. These compensatory mutations may only be selected for because of their epistatic effects in reducing the negative phenotypes of other mutations (de Visser, 2011).

When the effect of a mutation's fitness switches, depending on the background it arises on, this is termed as sign epistasis. A mutation may be beneficial on one genetic background and deleterious on another (Weinreich, 2010). For example, sign epistasis occurs between mutations that provide cefotaxime resistance in the B-lactamase enzyme (Weinreich, 2010). G238S and M182T are two mutations that are commonly found. On a wild type background, G238S is a mutation that increases cefotaxime resistance, but reduces the thermodynamic stability of the enzyme (Weinreich, 2010). On a wild type background, M182T increases the thermodynamic stability of the B-lactamase enzyme, but decreases the cefotaxime resistance and catalytic activity (Weinreich, 2010). When both mutations are present, a greater increase in drug resistance occurs than when compared to only the G238S mutation, even though on its own M182T decreases resistance. This indicates an epistatic interaction, as the sum of

the total antibiotic resistance is greater than the separate mutations provide on their own. More specifically, this is an example of sign epistasis, as the benefit of the M182T mutation depends on the background that it arises on: a wild type background would make it deleterious while an antibiotic resistant background would make it beneficial (Weinreich, 2010). Sign epistasis is also an example of how compensatory mutations can arise to offset the costs that sometimes come with beneficial mutations. Mutations may be neutral or deleterious when selected on a wild type background, but may compensate reductions in fitness that may occur because of antibiotic resistance. Often these fitness defects are seen as reductions of growth when the resistant microbial population is growing in the absence of antibiotics, with compensatory mutations arising to offset this fitness cost.

Finally, the interaction of phenotypes and the environment may also shape epistatic interactions, especially negative interactions. In an environment with few resources, negative epistatic interactions may be amplified, causing increased fitness costs. Additionally, there has been evidence that selection for epistatic interactions is increased in more complex environments (de Visser, 2011).

High-Throughput Screens

Synthetic Lethal/Synthetic Dosage Lethal Interactions

Understanding how genes interact epistatically can be determined using synthetic lethal (SL) and synthetic dosage lethal (SDL) interactions. These specific types of epistasis that have received attention in their value to researchers. Synthetic lethal

screens search for combinations of two non-allelic, non-essential mutations, that on their own have little or no negative effect on the organism, but when combined result in death (Kroll, 1996). Death implies an interaction is occurring between the two mutated genes of interest, because the other non-mutated gene masks or compensates the effect of the other mutation, but the combination of mutations together is lethal (Butland, 2008). The two mutations are often present in parallel non-essential pathways (Measday, 2005). Synthetic dosage lethal screens are similar, but instead of having two distinct mutations, a combination of a mutation and over-expression of a wild type allele causes lethality (Kroll, 1996). Although the over-expression of some genes can be lethal on their own, in yeast, only a few genes are toxic when overexpressed, so this is not considered a major problem (Kroll, 1996).

Such screens can identify epistatic interactions by looking at the impact of the mutations individually and then comparing them to the impact of the double mutants. If the effect on the double mutant (usually its effect on growth) is greater than the sum of the two mutations together, epistasis is inferred.

How Synthetic Lethal/Synthetic Dosage Lethal Interactions were previously identified

A variety of methods have been used to study SL and SDL interactions. A common tool in yeast to carry out systematic whole-genome analysis is the synthetic genetic array (SGA). SGA is a tool that searches for double mutants that cause a decrease in fitness, usually detected by a reduction of growth determined by colony size (Measday, 2005). Often a query mutation is created, based on the gene of interest.

Query mutations are often deletions, and strains with these deletions are then mated to a library of mutant strains with knockouts covering all non-lethal open reading frames, with a drug resistant marker in its place (Butland, 2008). Visual inspection or automated colony size measurements are then used to determine decreases in growth (Measday, 2005).

The first large screen used to detect synthetic lethal genetic interactions was implemented by Bender and Pringle, and was used to detect genes involved in morphogenesis (Bender, 1991). Using a colony-sectoring assay based on colour, they were able to identify mutants in which a morphogenesis gene *MSBI* became necessary for survival. They started with a deletion strain of *MSBI*, which appeared white in the colony-sectoring assay. Bender and Pringle then mated the deletion strain with another yeast strain containing autonomously replicating plasmids with a working copy of the *MSBI* gene, with genes causing the colonies to go red (Bender, 1991). UV mutagenesis was used to create secondary mutations in the genome, potentially hitting genes of interest. After mating, maintenance of the plasmid could be determined by observing completely red colonies (Bender, 1991). Colonies that were sectored and partially white indicated that the plasmid was lost. By searching for completely red colonies, this indicated that *MSBI* on the plasmid was necessary for survival, and a mutation was present in a gene of interest (Bender, 1991). This method relied on both visual characterization of the red phenotype, and luck, as it relied on UV mutagenesis to create a non-functioning gene of interest.

SGA studies uses arrays of non-essential deletion mutants and conditional mutants of essential alleles, handled by automated robotics to create haploid double mutants. These double mutants are then scored based on growth in either liquid using optical density or growth on plates using colony size (Costanzo, 2011).

In *E. coli* SGA is also used, and involves conjugation, which is similar to using mating in yeast. *E. coli* SGA uses the Hfr conjugation system and mutants from two deletion libraries that extensively cover the *E. coli* genome (Typas, 2008). Conjugation occurs between donor strains with query mutations and a library of recipient single gene mutants to create double mutants, which are then selected for using markers.

In both *E. coli* and yeast studies, double mutant colonies are quantitatively analyzed using an imaging program for differences in growth (Memarian, 2007). Growth defects suggest the presence of epistatic interactions. Some screens take advantage of searching for a positive interaction as well by looking for colonies that are larger than average (Typas, 2008).

The study of synthetic lethal interactions in bacteria is not common in literature, with only 200 lethal interactions being found in *E. coli* (Typas, 2008). Eukaryotes such as yeast are easier to work with, with established methods using mating to create double mutants. Additionally, in the past, the lack of fully annotated genomes for many species of bacteria, even when studying *E. coli*, made them less favourable to work with (Typas, 2008).

Despite these screens being considered as high-throughput, as they make the possibility of gaining extensive knowledge about many gene interactions possible, they

still take significant time and effort, and expensive equipment to create double mutants and visibly evaluate growth to determine synthetic lethality. These techniques are scored visually. Automated computer programs are now able to give colony growth or optical density a quantitative score, which is able to somewhat give a range of fitness effects rather than a yes or no answer to indicate an interaction. Subtle differences in growth and therefore fitness may not be detected by such analyses, resulting in a lack of ability to detect modest epistatic interactions.

Distribution of Epistatic Interactions determined from previous screens

Work from Costanzo et al (2010) indicates that epistasis in budding yeast is extensive, with up to 100,000 negative epistatic interactions being detected between 4,000 genes, with a median of 22 unique epistatic interactions per gene (Costanzo, 2011). Between 5.4 million gene-gene pairs, approximately 170,000 gene interactions were detected (Costanzo, 2010). Genes that are in the same pathway or that have similar biological functions tend to have the same type of genetic interaction, with genetic interactions between pathways either being all positive or all negative for when looking at pair wise interactions between genes (Costanzo, 2010). Synthetic lethal interactions have led to mapping complex systems, shedding insight on how the cell functions (Costanzo, 2010).

It has also been found that the more genetic interactions a gene has, the more tightly controlled the expression of that gene is. As such, more epistatic interactions

may be found by focusing on genes of high importance, such as those that interact with DNA(Park, 2013).

From an evolutionary point of view, it has been found that genes with many epistatic interactions are conserved across species (Park, 2013). The more interactions a gene has, the more likely there is a strong selection in maintaining stable expression of that gene, resulting in slower divergence and possibly constrained evolution of expression for those genes (Park, 2013).

Using Next Generation Sequencing to characterize Synthetic Dosage Lethal Interactions

One of the difficulties of studying synthetic lethal/synthetic dosage lethal interactions is that it is necessary to analyze each interaction individually, to determine if it is lethal. Each double mutant must be created and plated out separately, such that thousands of crosses are required for even a modest survey of genetic interactions. I propose that, by taking advantage of next-generation sequencing and whole genome plasmid libraries, a query gene can be analyzed across a whole genome's worth of over-expression plasmids in one experiment.

Next generation sequencing, unlike conventional Sanger sequencing, can sequence many samples at once. Although one of the main applications of next generation sequencings is to determine the sequence of the genome in question, because it gives many reads and those reads can be counted, it can be applied to ask quantitative questions, such as differences in transcriptional levels from RNA samples (Morozova, 2008). Using this quantitative feature, deep coverage can be obtained, and

using a reference genome, these reads can be mapped. From the sequencing information, both the presence of the RNA and the quantity are measured. Transcriptomes of cells exposed to different conditions or closely related species can be measured, and changes in these transcriptomes can be detected using this method (Morozova, 2008).

This way of quantifying reads can be further applied to measuring fitness of a bacterial cell in a population, as long as that strain can be identified in some way. By monitoring changes in frequency of different genotypes, we can then estimate fitness from these changes in frequency. These changes in frequency represent the selection coefficient, which is the slope in the change between two $\log_2$ time points over time (Hietpas, 2011).

Evolution and Epistasis

While epistasis can be studied from a purely functional point of view to better understand the genetic system of a cell, the underlying impact of epistasis on evolution often goes unnoticed. Major unanswered questions in understanding adaptation include the distribution of epistasis across systems, and how epistasis affects an evolving population (Orr, 2005).

Fisher and Wright, foundational population geneticists whose evolutionary models still shape theory today, had contrasting views on epistasis. Fisher believed that interacting loci were significant to evolution only if they are simultaneously polymorphic, leading to “statistical epistasis”. Fisher also believed it did not contribute

to genetic variance. Wright, rather than just looking at epistasis statistically across a population, was interested in the effect of a novel mutation on polymorphic background, and how that effect might differ, which can be differentiated from Fisher as “compositional epistasis” (Weinreich, 2005).

Wright introduced the concept that fitness, influenced by epistatic interactions, could be described as a landscape to explain adaptation and the process of natural selection in 1932 (Weinreich, 2005) (Wright, 1932). He thought populations could get stuck on local fitness peaks because there was no accessible trajectory to make it to the next peak, because due to epistasis only deleterious mutations on that background could exist (Weinreich, 2005). Despite this, epistasis was never truly mathematically considered in evolutionary models beyond specific cases such as reproductive isolation and canalization (Hansen, 2013).

Expanding on molecular trajectories, epistasis is an important consideration in looking at evolution and population genetics because of its influence on creating multiple fitness peaks (Clark, 1997). Multiple fitness peaks may prevent organisms from reaching their maximum fitness but may also influence speciation and contribute to the evolution of sex. Different genetic sequences inherently will have different fitnesses, and can be thought of as inhabiting different places on a fitness landscape where the axes are genotypes and the height is fitness. By plotting combinations of genotypes at different loci different combinations will produce different fitness peaks (Orr, 2005). Fitness landscapes can either look like one major peak, with the populations slowly moving towards one optimum, or have many different optima, with a variety of sizes

(Orr, 2005). Once a population reaches the top of a fitness peak, it may have no way to make it to an adjacent peak, even if that peak is at a higher fitness, because it first would have to become less fit to get there. Only if the environment changes, and therefore the peaks shift, would new mutations be selected for (Weinreich, 2005).

Tenaillon and colleagues conducted an experiment where 115 populations of *E. coli* were evolved for 2000 generations. They found a clear effect of epistasis on the trajectory of evolution. Populations selected under high temperatures adopted either one of two paths, depending on the identity of the first mutation to fix. Adaptive mutations that arose made other possible adaptive mutations neutral or deleterious. It was shown that epistasis was prevalent in shaping the adaptive trajectories of the populations (Tenaillon, 2012).

Regardless of if statistical, functional, or compositional epistasis was deemed important, mathematical population genetics theory is still based in the idea that a gene has the same effects regardless of background, largely ignoring epistasis altogether (Hansen, 2013). Despite large SGA studies done with yeast showing the extent of epistatic interactions being widespread, often little effort has been made to include epistasis in evolutionary theory. Some researchers (e.g., Wade 2002) have studied the impact of epistasis on speciation and population subdivision, but often epistasis is only included in very specific models such as gene canalisation (Wade, 2002).

Antibiotic Resistance

Antibiotic resistance arises from the selection of randomly arising mutations in a population of bacteria that confer resistance to an antibiotic. Since the basis of antibiotic resistance arises at the DNA level, understanding epistatic interactions as a result of new mutations may provide us with some insight on future secondary mutations that either offset the costs of antibiotic resistance, or reverse the effects of antibiotic resistance.

Widespread resistance/Impact on Community

Antibiotic resistance or antimicrobial resistance (AMR) is an internationally growing problem, resulting in hundreds of thousands of deaths annually (Palmer, 2013). The 2014 global report on AMR surveillance, released by the World Health Organization (WHO), describes the prospect of a post-antibiotic future, where currently treatable infections become untreatable, as a very real risk, rather than an “apocalyptic fantasy” (World Health Organization, 2014). High levels and frequency of resistance have been found in bacteria that are clinically significant, in that they commonly cause infections, such as *Escherichia coli*, *Klebsiella pneumoniae* and *Staphylococcus aureus* (World Health Organization, 2014).

Common examples of antibacterial treatment failing include the inability to use penicillin to treat pneumonia caused by infections, common neonatal infections, and life threatening infections of immunosuppressed patients such as those receiving cancer treatment or undergone organ transplants (World Health Organization, 2014). Such

examples show how antibiotic resistance has an impact on the community and patients with these infections.

The report's solution emphasized surveillance, both of human and animal use, to be able to provide data to create new policies and strategies to reduce the risk of the coming post-antibiotic era. While surveillance and monitoring of AMR is important, the need for new antimicrobials is a problem that simply will not go away with surveillance. Unfortunately the last new classes of antibacterial drugs were discovered in the 1980s (Silver, 2011).

Since the 1980's, there has not been a lack of innovation in screening for new compounds, but rather a withdrawing of industry and pharmaceutical companies to invest in research for new antibacterials (Silver, 2011). In screens to find compounds that kill bacteria, it is easy to find compounds, but often they are not novel compounds and have been previously discovered. Finding the same compounds consistently over and over make it difficult in finding new compounds to develop (Silver, 2011). Most of our antibiotics come from actinomycetes, a class of soil bacteria that has led to our knowledge and production of streptomycin and vancomycin (Clardy, 2009) and resulted from empirical screening (Silver, 2011). In the last 30 years, instead of finding new classes of compounds, modifications have been made to known old ones, in order to maintain the compounds use and avoid resistance (Silver, 2011).

In the past, little concern has been given to the mechanism of action of a compound. While this has previously worked, with continued screening the same natural compounds have come up over and over again, with very few new classes of

antibiotic compounds being found (Clardy, 2009). An alternative strategy is to start with a gene target, and work in reverse to find a molecule that affects that target in the desired way. The difficulty is in finding those gene targets, in which thousands of combinations may be screened. By using a high-throughput epistatic screen, we can search for gene targets, which kill only antibiotic resistance strains. This is beneficial as having drugs that target only antibiotic resistant strains will provide treatment options for antibiotic resistant infections, but could also be used in combination with other antimicrobials, so if resistance arises, it can be combatted with the resistance targeted drug.

Resistance in E. coli

Antimicrobial resistance in *E. coli* is widespread. *E. coli* is often implicated in urinary tract infections, is the most frequent cause of sepsis, and one of the top causes of food borne infections world wide (World Health Organization, 2014).

Fluoroquinolones used to be the preferred antibiotic for treating *E. coli* infections, especially for urinary tract infections, but resistance is now prevalent (Johnson, 2013). In developing countries quinolones may be the last available oral treatment, with the next option being injections, meaning a greater cost to treat such infections (World Health Organization, 2014).

Fluoroquinolones function by preventing DNA synthesis from occurring. Quinolones inhibit DNA replication in bacteria by targeting DNA binding sites of DNA gyrase and DNA topoisomerase (Chen and Lo, 2001). DNA gyrase is made of two subunits, encoded by the genes *gyrA* and *gyrB* (Chen and Lo, 2001). Gyrase is important

to bacterial replication because it removes positive supercoils during DNA replication and adds negative supercoils following replication, allowing the DNA to be tightly packed into the cell. Quinolones interfere by binding to the gyrase subunits while the DNA is cleaved, and blocking DNA replication, which eventually leads to cell death (Chen and Lo, 2001).

Ciprofloxacin is a type of fluoroquinolone that is typically used to treat *E. coli* infections. Increasing prevalence of resistance has been observed worldwide, with up to half of invasive isolates or isolates causing infections such as meningitis or sepsis being resistant to fluoroquinolones (World Health Organization, 2014). Resistance can arise via numerous mutations, with the most common being a point mutation resulting in a change of Serine to Leucine at the 83rd amino acid in the GyrA protein (referred to as S83L) (Gales, 199). This mutation prevents ciprofloxacin from binding to GyrA, preventing it from stopping supercoiling, and DNA synthesis continues. Resistance mutations can occur in other genes such as *gyrB*, *parC*, and genes involved in efflux pumps, but the S83L mutation arises most frequently in resistant isolates (Uchida, 2010).

Using epistasis- finding new drug targets through a high-throughput screen

Epistatic interactions can be used to find new drug targets by searching for synthetic dosage lethal/sick interactions on antibiotic resistant background strains in comparison to wild type strains. Looking for synthetic dosage lethal/sick interactions would allow targeting of resistant strains only. In our screen, we are using the S83L *gyrA* antibiotic mutant background

and searching for overexpressed genes that cause a synthetic dosage lethal or sick interaction once induced. These genes could then be used as potential candidates for drug targets and drugs could be designed to mimic their action or cause over-expression.

By looking for a difference in selection coefficients determined from overexpression plasmid ORF frequencies changing over time, on both the mutant and wild type background, we will be able to infer epistasis. Additionally, we will be able to describe the distribution of epistatic effects for the *gyrA* antibiotic resistant point mutation, for both beneficial and deleterious interactions which not only has implications medically, but also from an evolutionary point of view. If epistatic interactions are widespread, evolutionary models may need to be therefore adjusted.

Materials and Methods

Construction of the quinolone resistant mutant genotype

Construction of the gyrA S83L allele

Genomic DNA was extracted from *E. coli* MG1655 using the Promega Wizard Genomic DNA Isolation Kit. I used the megaprimer method of the Herring lab to construct the S83L allele of *gyrA* (Herring, 2003). In megaprimer PCR, two steps are used to generate an allele. In the first step, two primers are used to amplify the first half of the allele, from the beginning of the gene up until the desired point mutation. The reverse primer is designed to contain the wanted point mutation, which then produces half of the allele, with the desired point mutation near the end of the PCR product. In the second step, the PCR product of the first reaction is used at the first primer, dubbed

the “megaprimer”, and a second primer to finish the allele. The product of this reaction is now the complete mutant allele, which I then used in gene gorging (Herring, 2003).

PCR with the genomic DNA, using the following primers mkGyrA-F 5'-ATGAGCGACCTTGCGAGAG-3' and mkGyrA-R1-S83L 5'-CGTGTCATAGACCGCCAAGTCACCATGGGGATG-3' was used to create the megaprimer, and to introduce a point mutation in the *gyrA* gene, at the 83rd codon, which had the nucleotides TCG, changing it to TTG. This mutation changed the amino acid from serine to leucine.

Using the unpurified Megaprimer and the following primer mkGyrA-R2-Sce 5'-ATTACCCTGTTATCCCTATTTTCGAGATCGGCCATCAG-3', PCR was performed with the isolated genomic DNA of MG1655. This was done to finish the gene and add in a restriction site 1-*SceI*, necessary for gene gorging. Unpurified megaprimer was used, as purified megaprimer did not yield any PCR product.

Using the TOPO Cloning kit, the second PCR product was cloned into TOP 10 PCR2.1 with pCR Vector Plasmids were extracted using a Wizard miniprep isolation kit. PCR with the cloned plasmids was performed with primer mkGyrA-F-M13F 5'-TGTAACGACGGCCAGTATGAGGCACCTTGCGAGAG-3' and mkGyr-R2-Sce was done to add the M13 primer site for sequencing. Sequences were sequenced at Genome Quebec to ensure mutation and restriction site was present, as well as no other mutations. This plasmid was then used in gene gorging as the donor plasmid.

Gene gorging

The gene gorging method of the Herring Lab was used to induce recombination of our *gyrA* S83L mutant allele, which was the megaprimer PCR product, into the wild type *E. coli* genome (Herring, 2003). Gene gorging consists of transforming a recipient strain with two plasmids: the donor plasmid containing the mutant allele, which in this case contains the mutant *gyrA* product of the Megaprimer PCR, and the mutagenesis plasmid (pACBSR low copy plasmid). The donor plasmid allele must also contain the restriction site 1-SceI, used in the recombination process.

The mutagenesis plasmid is induced by L-arabinose, which turns on two sets of genes needed for recombination of the donor plasmid with the wild type genome: a meganuclease that cuts at the 1-SceI site, which then linearizes the donor plasmid, and the lambda red recombination genes that direct recombination. These genes are encoded on pACBSR. Once the donor plasmid is linearized, it provides the substrate for lambda red gene products, and recombination occurs at the matching ORF, replacing the wild type allele.

Chemically competent cells of MG1655 *E. coli* were created using the Kantz Lab protocol. Using protocol from the Barrick lab, chemically competent cells were transformed with 1 ul of donor *gyrA mutant* plasmid and 1 ul of Mutagenesis plasmid (pACBSR low copy plasmid) (Herring, 2003). Chemically competent cells were heat shocked at 42 C for 30 seconds. Cells were then grown with 500 ul SOC liquid medium consisting of for one hour at 37 °C in a shaking incubator to induce antibiotic expression. Cells were then grown on LB agar plates consisting of 10 g/L tryptone, 5 g/L yeast

extract, 10 g/L sodium chloride, and 15 g/L agar, with 25 ug/ml chloramphenicol and 50 ug/ml kanamycin, to ensure double transformation.

Colonies were picked from these plates and grown overnight at 37 °C in LB supplemented with 20 ug/ml chloramphenicol, 50 ug/ml kanamycin and 0.2% L-Arabinose to induce gene gorging. The original protocol did not include kanamycin in the liquid culture when inducing gene gorging. Adding kanamycin was found to increase the efficiency of the protocol, because it selected for the gene gorging plasmid. Liquid culture was then plated first to LB and incubated overnight at 37 °C. Colonies were then replica plated on plates that contained either with 25 ng/ul ciprofloxacin, 25 ug/ml chloramphenicol, 50 ug/ml kanamycin and both chloramphenicol and kanamycin, to show antibiotic resistance but also plasmid loss, to identify colonies in which gene gorging had occurred. Colonies that were lethal on the chloramphenicol and kanamycin, but survived on the ciprofloxacin were candidates. Freezer stocks were made of potential candidates. We did not select colonies straight from the ciprofloxacin plates to avoid selection for other mutations leading to antibiotic resistance.

DNA was extracted from potential gene gorging candidates, *gyrA* was amplified by PCR, and the PCR product was sequenced at McGill University and Genome Quebec Innovation Centre, to confirm the presence of the S83L mutation.

Sequencing of a PCR reaction of the *gyrA* gene from the newly made mutant genome after gene gorging (Figure 2) confirmed the presence of the S83L point mutation, from a T to C point mutation at the 248th nucleotide, without additional mutations in *gyrA*. Additionally, the presence of the point mutation in the mutant

genome was confirmed by performing an MIC assay to prove that the strain was now resistant to ciprofloxacin. MIC values for the mutant and wild type were determined by Liam O’Sullivan. The mutant’s MIC90 value is 1000 ng/ml, where the wild type MIC50 values is 40 ng/ml, indicating that the S83L point mutation results in at least a 25-fold increase in Ciprofloxacin resistance.

Score	Expect	Identities	Gaps	Strand	
682 bits(369)	0.0	371/372(99%)	0/372(0%)	Plus/Plus	
Query 1	CTGAAGAGCTCCTATCTGGATTATGCGATGTCGGTCATTGTTGGCCGTGCGCTGCCAGAT				60
Sbjct 49	CTGAAGAGCTCCTATCTGGATTATGCGATGTCGGTCATTGTTGGCCGTGCGCTGCCAGAT				108
Query 61	GTCCGAGATGGCCTGAAGCCGGTACACCGTCGCGTACTTTACGCCATGAACGTACTAGGC				120
Sbjct 109	GTCCGAGATGGCCTGAAGCCGGTACACCGTCGCGTACTTTACGCCATGAACGTACTAGGC				168
Query 121	AATGACTGGAACAAAGCCTATAAAAAATCTGCCCGTGTTCGTTGGTGACGTAATCGGTAAA				180
Sbjct 169	AATGACTGGAACAAAGCCTATAAAAAATCTGCCCGTGTTCGTTGGTGACGTAATCGGTAAA				228
Query 181	TACCATCCCCATGGTGACTTGGCGGTCTATGACACGATCGTCCGCATGGCGCAGCCATTC				240
Sbjct 229	TACCATCCCCATGGTGACTCGGCGGTCTATGACACGATCGTCCGCATGGCGCAGCCATTC				288
Query 241	TCGCTGCGTTATATGCTGGTAGACGGTCAGGGTAACTTCGGTTCTATCGACGGCGACTCT				300
Sbjct 289	TCGCTGCGTTATATGCTGGTAGACGGTCAGGGTAACTTCGGTTCTATCGACGGCGACTCT				348
Query 301	GCGGCGGCAATGCGTTATACGGAAATCCGTCTGGCGAAAATTGCCCATGAACTGATGGCC				360
Sbjct 349	GCGGCGGCAATGCGTTATACGGAAATCCGTCTGGCGAAAATTGCCCATGAACTGATGGCC				408
Query 361	GATCTCGAAAAA	372			
Sbjct 409	GATCTCGAAAAA	420			

Figure 2. The aligned sequences from the *gyrA* gene. The query strand is the PCR product from the S83L *gyrA* mutant genome after gene gorging. The subject strand is the MG1655 wild type *gyrA* gene. There are no differences between the two except for a C to T mutation at the 248th nucleotide.

Plasmid Library Replication

Both GFP positive and negative ASKA over-expression plasmid libraries for *E. coli* were sent to us from Nara Institute of Science and Technology, Japan. The ASKA over-

expression plasmid library is based on the genome of *E. coli* K-12 and represents the complete set of cloned ORFs (Kitagawa, 2006). Each of the 4123 different strains contains one different plasmid coding an ORF, with an *Sfi*1 restriction site, and genes for chloramphenicol resistance (Kitagawa, 2006). Gene expression is induced by an IPTG-inducible promoter. The library comes both GFP tagged and non-tagged (Kitagawa, 2006).

The ASKA library was received on agar plates and were made into freezer stocks. Using a handheld pinning tool both the GFP positive and negative stocks from the agar plates were transferred to 96 deep well plates (twice for the GFP negative library) filled with liquid LB and 30 ug/ml chloramphenicol and incubated overnight at 37 °C in a shaking incubator. Equal amounts of 50% glycerol were added to the wells and then kept at -80 °C.

Plasmid Library Extraction

Using the Promega Wizard Midiprep kit, plasmids from the GFP negative ASKA over-expression library were extracted for each ORF (over 4000 different strains). Strains were grown overnight in 96 well plates with 30 ug/ml chloramphenicol and LB at 37 °C in a shaking incubator. Two 96 well plates were pooled together for each midiprep to reduce extraction numbers.

The extracted plasmids were mixed together to make a complete solution containing all 4000+ over-expression vectors. Concentrations were equalized from each

pool of 192 plasmids, so that roughly the same amounts of plasmids were added from each batch to the complete mix.

High-throughput screen for genetic interactions

The high-throughput screen is described in Figure 3. Chemically competent *gyrA* S83L mutants and wild types (MG1655) were produced using Berkeley's Protocol. Three tubes of each were used in the transformation. 3 ul of plasmid solution containing all 4000+ ORFs from the ASKA over-expression library were added to each tube, which was then mixed by flicking, and was allowed to sit on ice for 30 minutes. Each tube was then heat shocked at 42 °C for 1 minute. Cells were then grown with 500 ul SOC medium for one hour at 37 °C in a shaking incubator. For each strain the three tubes were combined and added to 50 ml of LB liquid broth and 30 ug/ml chloramphenicol and incubated for 3 hours.

10 ml of the 3 hour long incubation culture were then added to 1 L of liquid LB with 30 ug/ml chloramphenicol, 1mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) to induce plasmid expression and 4 different concentrations of ciprofloxacin: 0 ng/ml, 5 ng/ml, 10 ng/ml and 50 ng/ml for each strain, giving a total of eight conditions. 10 ml of culture were reserved, and 10 ml of 50% glycerol were added and frozen, representing the time 0 (t-0) sample. The resulting cultures were incubated for 24 hours at 37 °C.

20 ml of each sample condition were then taken every 30 minutes for the first 6 hours. 20 ml of each sample condition were then taken every hour for the next 6 hours,

and once more after 24 hours. 20 ml of 50% glycerol were added to each sample and frozen at minus 80 °C. Three replicates of the experiment were performed.

Figure 3. Visual description of high-throughput screen, with samples used in sequencing shown.

Pre-Sequencing Preparation

Plasmids from each replicate were extracted from the MG1655 and S83L mutant populations for concentrations 0 ng/ml and 10 ng/ml ciprofloxacin at the following time points: t-0, 2, 4, 6, 8, 12, 24 hours. Plasmids were extracted from 20 ml of thawed frozen

sample using the Bio Basic EZ-10 Spin Column Plasmid DNA Kit, using their low copy plasmid protocol.

Sequencing

Quantification of the plasmid frequencies from the t-0, -6, and -24 hour samples was performed using the Illumina Miseq Benchtop Sequencer (2 x 300-bp paired end sequencing). We used the Nextera-XT kit to prepare sample libraries, and loaded 9 samples per cartridge. Our read count was on average 2743028, in one direction. Since we obtained paired-end reads, on average 653069.5 reads per sample aligned to the reference genome.

Bioinformatics and Data Analysis

Bioinformatics workflow

Trimmomatic was used to remove the Illumina sequencing adapters, which was built into the program (Figure 4). The command used to specify which adapters was the TruSeq3-PE command. The parameters used were remove leading or trailing low quality bases below 3, sliding window of 4:15, so when the average quality of a 4 base window went below 15, the bases were trimmed, and it was also specified to drop reads less than 36 bases long.

Trimmed paired and unpaired end files were then combined back into two files, to have all information from one end in one file.

Reads were converted from fastx to fasta format using a perl script. This was because Seqclean requires the reads to be in fasta format. Seqclean was used to trim the vector sequences. The vector sequence on average made up half of the reads.

Bowtie was used to align the reads to a reference genome of MG1655. The SAM file was then used in HTSeq, to count the reads, to determine the frequency of each gene and therefore each plasmid at each timepoint.

Figure 4. Bioinformatics workflow used to identify the genes, and the quantity present in the extracted plasmid.

Data and Statistical analysis

Once gene counts were determined, the frequency of each gene was calculated, by dividing the individual gene count by the total gene count for that timepoint. To adjust for genes that had a frequency of 0, 0.0000001 was added to each frequency, allowing a selection coefficient to be calculated (see below). The selection coefficient, which was determined by finding the rate of change over time of each for each gene was calculated by taking the natural log of the frequency at one time point, and subtracting from it the natural log of the frequency at the starting time point, while dividing by the time that had passed.

$$S_{\text{gene } i} = \frac{(\log(\text{gene}_i \text{ frequency at 6 hours}) - \log(\text{gene}_i \text{ frequency at 0 hours}))}{6 \text{ hours}}$$

The selection coefficient was then normalized for the effects of general selection and competition in the assay by subtracting the median pseudogene selection coefficient determined individually for each sample from 4 pseudogenes, from each selection coefficient in the given sample. This was done because it is expected that the pseudogenes had no epistatic effect in regards to the mutation, but generally experienced a decrease in frequency over time. By subtracting the median pseudogene selection coefficient from samples, we could more readily accept that the resulting epsilon value calculated was a true reflection of any gene interactions occurring, and not just from competition.

$$E = S_{\text{gene X for WT}} - S_{\text{gene x for MT}}$$

The normalized selection coefficient was then used to infer epistasis. Selection coefficients were averaged over the three trials for each gene. It was found that trial 3's selection coefficients for both the wild type and the mutant did not match trials 1 and 2. The average selection coefficient was then calculated for trials 1 and 2 for each gene, which was used to calculate epsilon. Epsilon, representing epistasis, was calculated by subtracting the average selection coefficient of the wild type from the mutant selection coefficient for each gene. Epistasis was inferred with values farther from zero, where as a positive value indicating epistasis was beneficial for the wild type, and a negative value indicating epistasis was beneficial for the mutant (Bell, 1997).

For each over-expression construct, statistical significance of differences in selection coefficient between the wild-type and the mutant background was calculated using a binomial test. For a given gene with frequency f_0 at time 0 and selection coefficient s , the expected frequency at 6 hours is given by

$$E(f_6) = e^{6s + \ln(f_0)}$$

The probability of the null hypothesis, that s is the same for both the wild-type and mutant background, was calculated using the `binom.test()` function in R, given the observed counts for a given gene, the total number of counts across all genes, and $E(f_6)$ as the probability of success. Epsilon values were considered significant when paired with P -values of less than 0.0001 for both trials.

Additionally, 24 genes were removed from the data set because read counts were too low. The criterion for removal was at least one trial having a starting frequency of less than 10.

Epistatic interaction validation

Potential epistatic interactions were confirmed by transforming the plasmid bearing the gene of interest individually into both the wild type and mutant backgrounds. First plasmids strains were grown overnight at 37 °C in 30 ug/ml chloramphenicol, and using Bio Basic EZ-10 Spin Column Plasmid DNA Kit. Transformations were performed with chemically competent MG1655 and S83L mutant cells created using the Kantz Lab protocol, in the same manner as done in the high-throughput experiment. After the transformation, cells were streaked for single colonies onto plates and grown at 37 °C overnight on LB agar plates with 30 ug/ml chloramphenicol. 2 ul of liquid culture was then added to 96 well plates containing 198 ul LB broth, 30 ug/ml chloramphenicol and 0.1 mM IPTG in triplicate, and was added to 1 well with only LB broth and 30 ug/ml chloramphenicol. 24 hour growth curves were collected, measuring OD600 every 15 minutes using a BioTek ELx808 plate reader. A difference in growth curves, determined by looking at maximum growth rate and optical density after 6 and 24 hours, between the wild type and mutant indicated epistatic interactions and confirmed results from the high throughput screen. Differences

between the mutant and wild type were detected performing a two-factor ANOVA with statistically significant results have P -values less than 0.05.

Results

High-throughput screen for genetic interactions

Sequencing and bioinformatics workflow

The initial sequence counts for each sample, the sequences kept after trimming with Trimmomatic, the sequences kept, trimmed and discarded due to vector removal by SeqClean, and aligned to the reference genome using Bowtie were determined (Table 1). Trial 1 for the wild type and mutant, and trial 2 for the wild type were sequenced in the Carrillo laboratory at the Canadian Food Inspection Agency. Trial 3 for the wild type and mutant and trial 2 for the mutant were sequenced in the Wong lab. CFIA's Miseq software does not trim the adapters, while the Wong Lab sequencer does, resulting in a difference between Trimmomatic results between those trials. Up to 90% of the sequences for each sample were lost due to discarding of the vector sequences.

Table 1: The reads for each sample after being sequenced by an Illumina Miseq sequencer, the sequences kept after being trimmed with Trimmomatic, the sequences kept, trimmed and discarded due to vector removal by SeqClean, and final count of sequences aligned to the reference genome using Bowtie.

Sample ^a	Reads	Kept After	Kept after Seqclean	Trashed	Total Aligned to
---------------------	-------	------------	---------------------	---------	------------------

	Trimmomatic			after Seqclean	Reference Genome ^b
t0wt1R1	1113688	1110449	1041639 (9642 trimmed)	68810	893891
t0wt1R2	1113688	1110439	1041629 (9642 trimmed)	68810	
t6wt1R1	2219455	2772616	410913 (103476 trimmed)	2361703	257882
t6wt1R2	2219455	2635147	374037 (86554 trimmed)	2261110	
t24wt1R1	3441570	2456708	302698 (141205 trimmed)	2154010	258915
t24wt1R2	3441570	2279800	266041 (118656 trimmed)	2013759	
t0wt2R1	723212	716488	540640 (3733 trimmed)	175848	424333
t0wt2R2	723212	716486	540638 (3733 trimmed)	175848	
t6wt2R1	2473991	2449185	438893 (73621 trimmed)	2010292	359889
t6wt2R2	2473991	2311689	394105 (59277 trimmed)	1917584	
t24wt2R1	4524631	4454688	344523 (114874 trimmed)	4110165	311476
t24wt2R2	4524631	4225497	309453 (95103 trimmed)	3916044	
t0wt3R1	2123227	2123227	1937358 (8232 trimmed)	185869	1574105
t0wt3R2	1888810	1888810	1655782 (5385 trimmed)	233028	
t6wt3R1	2625487	2625487	1782283 (29182 trimmed)	843204	1482099
t6wt3R2	2388144	2388144	1546238 (21330 trimmed)	841906	
t24wt3R1	8402668	8402668	1132454 (222313 trimmed)	7270214	991656
t24wt3R2	7824377	7824377	998316 (173389 trimmed)	6826061	
t0mt1R1	739449	736365	634611 (10139 trimmed)	101754	475956
t0mt1R2	739449	736361	634607 (10139 trimmed)	101754	
t6mt1R1	2785740	2772616	410913 (103476 trimmed)	2361703	338253
t6mt1R2	2785740	2635147	374037 (86554 trimmed)	2261110	
t24mt1R 1	2463239	2456708	302698 (141205 trimmed)	2154010	272834
t24mt1R 2	2463239	2279800	266041 (118656 trimmed)	2013759	
t0mt2R1	2826428	2826428	1546799 (22832 trimmed)	1279629	1325208

t0mt2R2	2641897	2641897	1389013 (18014 trimmed)	1252884	
t6mt2R1	2253164	2253164	395059 (120575 trimmed)	1858105	314656
t6mt2R2	1948027	1948027	314054 (80708 trimmed)	1633973	
t24mt2R 1	2104459	2104459	226975 (104778 trimmed)	1877484	191034
t24mt2R 2	1848195	1848195	173872 (67545 trimmed)	1674323	
t0mt3R1	1181381	1181381	1026379 (5308 trimmed)	155002	828269
t0mt3R2	1055137	1055137	875046 (3044 trimmed)	180091	
t6mt3R1	517265	517265	442221 (5927 trimmed)	75044	319917
t6mt3R2	443302	443302	364234 (2897 trimmed)	79068	
t24mt3R 1	8165004	8165004	1306016 (223100 trimmed)	6858988	1134878
t24mt3R 2	7542096	7542096	1142789 (170952 trimmed)	6399307	

^aSample designated by the time in hours (0, 6, 24), mutant (mt) or wild type (wt), trial number (1-3) and Read number (1 or 2)

Sample of sequence counts using HTSeq

An example of the gene counts given for the first 8 genes for both paired-end reads, for all samples is found in Table 2. Over time, gene counts generally decrease, often decreasing to 0. Table 3 also shows the variation in initial gene counts at t-0. Initial counts ranged from 1,000 copies of the gene present, to only 1 or 2 copies.

Table 2: The raw gene counts for each sample, determined by aligning reads to the reference genome. The top panel shows the gene counts for the wild type, and the bottom panel shows the gene counts for the S83L *gyrA* mutant.

	t0wtJR1	t0wtJR2	t6wtJR1	t6wtJR2	t24wtJR1	t24wtJR2	t0wt2R1	t0wt2R2	t6wt2R1	t6wt2R2	t24wt2R1	t24wt2R2	t0wt3R1	t0wt3R2	t6wt3R1	t6wt3R2	t24wt3R1	t24wt3R2
b0001	24	24	6	2	0	0	10	10	4	2	0	0	32	37	36	28	4	11
b0002	215	215	55	64	89	59	137	137	67	62	4	1	500	416	450	433	120	102
b0003	81	81	86	66	18	14	42	42	65	57	6	4	160	101	161	117	83	52
b0004	109	109	48	32	7	0	43	43	35	38	5	0	188	191	216	207	68	48
b0005	52	52	3	4	2	0	15	15	10	16	0	0	96	93	93	75	29	24
b0006	82	82	13	16	1	1	36	36	25	14	1	0	167	156	157	124	56	45
b0007	171	171	24	19	4	0	73	73	48	27	0	0	322	252	327	246	140	105
b0008	147	147	36	36	5	4	54	54	93	64	6	3	246	215	296	234	85	61

	t0mtJR1	t0mtJR2	t6mtJR1	t6mtJR2	t24mtJR1	t24mtJR2	t0mt2R1	t0mt2R2	t6mt2R1	t6mt2R2	t24mt2R1	t24mt2R2	t0mt3R1	t0mt3R2	t6mt3R1	t6mt3R2	t24mt3R1	t24mt3R2
b0001	17	17	3	2	0	0	24	19	2	0	0	0	36	27	8	6	10	7
b0002	166	166	67	68	4	0	452	375	64	50	4	0	266	217	123	89	196	174
b0003	52	52	86	61	6	5	166	127	44	20	0	1	72	49	41	19	83	70
b0004	53	53	76	70	8	4	232	175	27	23	0	0	109	91	47	45	108	90
b0005	24	24	6	6	0	0	84	70	6	2	0	0	40	39	17	21	28	36
b0006	45	45	25	19	1	0	128	154	19	12	0	0	104	71	53	29	50	56
b0007	113	113	31	24	3	1	270	215	37	20	1	1	169	148	76	57	140	90
b0008	67	67	26	26	1	0	234	195	40	39	1	0	136	103	68	44	94	107

Sample of data and statistical analysis

An example of the data analysis for the first 8 genes is provided (Table 3). Below is a sample calculation for gene b0001, *thrL*, to determine the selection coefficient for both the wild type and mutant strains and the epsilon value, representing epistasis. First the frequency of the gene count was determined for both t-0 and t-6 for hours, for the wild type and mutant by dividing the gene counts by the total gene counts from that sample:

$$\text{Frequency at 0 hours for wt trial 1} = \frac{\text{Gene count at 0 hours}}{\text{Total gene counts}} = \frac{48}{893891} = 5.36978\text{E-}05$$

See table 3, for the frequencies at 6 hours for the wild type and 0 and 6 hours for the mutant for trial 1.

0.0000001 was added to each frequency, to remove any zero frequencies, which would be problematic in the equation to determine the selection coefficient.

Selection coefficient between 0 – 6 hours for wt trial 1

$$\begin{aligned} &= \frac{(\ln(\text{frequency at 6 hours}) - \ln(\text{frequency at 0 hours}))}{6 \text{ hours}} \\ &= \frac{(\ln(0.00003112194) - \ln(0.000053797822))}{6 \text{ hours}} \\ &= -0.09122 / \text{hour} \end{aligned}$$

To take into account the general decline in frequency that most genes displayed and to ensure that the frequencies used in the calculations were reflection of the gene effects, and not selection effects created by the format of the experiment. If some genes are advantageous and increase in frequency, neutral genes may decrease in frequency

because they are being out competed rather than dying due to a synthetic dosage lethal epistatic interaction. By comparing each gene to neutral candidates pseudogenes, I hope to remove the effects of other forms of selection from my results. I subtracted the median selection coefficient determined from 4 pseudogenes (b3504, b3683, b4236, and b4462) from each sample. To do this, I determined the selection coefficient for each pseudogene using the equation above, and then determined the median pseudogene selection coefficient for each time window, for each trial. The median pseudogene selection coefficient was subtracted from all the selection coefficients in that time window and trial. Pseudogene median selection coefficients were generally close to zero with the exceptions of trial 2 for the wild type after 24 hours at -0.1686663, trial 1 for the mutant after 24 hours at -0.1653684 and trial 2 for the mutant after 24 hours at -0.1027961.

After being normalized, selection coefficients ranged from negative, indicating a decline in gene frequency over time, close to 0, indicating only a small change in gene frequencies over time, or positive, indicating an increase in gene frequency over time.

Once the normalized selection coefficient was calculated, it was used to find epsilon using the equation below.

$$\begin{aligned}
 \textit{Epsilon after 6 hours for trial 1} &= SC \textit{ WT after 6 h} - SC \textit{ MT after 6 h} \\
 &= -0.067368073 - -0.207091918 \\
 &= 0.139723844
 \end{aligned}$$

Table 3. Data analysis for the first 8 genes to show calculated frequencies, selection coefficients, and epsilon for the first trial at 6 hours.

	t0wt1	t6wt1	t0wt1f	t6wt1f	t0wt1f+0.0000001	t6wt1f+0.0000001	St6wt1	St6wt1-PG	E61
b0001	48	8	5.36978E-05	3.10219E-05	5.37978E-05	3.11219E-05	-0.09121999	-0.067368073	0.139723844
b0002	430	119	0.000481043	0.000461451	0.000481143	0.000461551	-0.006928527	0.01692339	0.054937728
b0003	162	152	0.00018123	0.000589417	0.00018133	0.000589517	0.196497321	0.220349238	0.050814184
b0004	218	80	0.000243878	0.000310219	0.000243978	0.000310319	0.040087607	0.063939524	-0.101286071
b0005	104	7	0.000116345	2.71442E-05	0.000116445	2.72442E-05	-0.242096817	-0.218244901	-0.099218929
b0006	164	29	0.000183468	0.000112455	0.000183568	0.000112555	-0.081524154	-0.057672237	-0.050309826
b0007	342	43	0.000382597	0.000166743	0.000382697	0.000166843	-0.138365114	-0.114513197	0.009032825
b0008	294	72	0.000328899	0.000279197	0.000328999	0.000279297	-0.027296302	-0.003444385	0.042374744

Evaluation of consistency between trials

There is a correlation of the selection coefficients between trials 1 and 2 but not for trial 3 for the wild type and the mutant (Figure 5). Trials 1 and 2 correlate well for both the wild type and the mutant, and follow a linear path. Trial 3 does not correlate well for either the wild type or the mutant for trials 1 and 2, prompting us to focus on trials 1 and 2, which were more consistent.

Figure 5. Selection coefficient correlations for a) wildtype and b) mutant between the 3 trials.

Potential epistatic interactions

The distribution of epsilon for all 4123 different genes at 6 hours was determined (Figure 6). Epsilon represents epistasis on the x-axis. Epsilon is centered on 0, showing that most genes displayed little epistasis. As we move away from 0, there is a large drop off in numbers of genes showing epistasis effects, with a larger amount showing low levels of epistatic effects deleterious for both the wild type (left side of 0) or mutant (right side of 0). The distribution shown is what was expected, with most genes showing no epistasis, some showing low levels, and a few showing high levels of epistatic interactions.

Figure 6: The distribution of epistatic interactions at 6 hours. Epsilon on the x-axis is the difference in fitness between the wild type MG1655 *E. coli* strain and the S83L *gyrA* mutant strain.

Fitness of each over-expression construct was highly correlated between the wild type and mutant backgrounds (Figure 7). Most genes do not show significant epistatic interactions and are displayed in black. Genes showing epistatic interactions, are in red. The dashed line is the 1:1 line, where fitness is equal on both backgrounds. Red dots above the dashed line are synthetic dosage sick/lethal for the wild type, and red dots below the dashed line are synthetic dosage sick/lethal for the mutant. 216 statistically significant candidates were detected, with 149 of them being negative for the mutant, and 67 being positive for the mutant.

Figure 7. Fitness for all 4123 over-expression genes on the MG1655 wild type background versus the *gyrA* S83L mutant background after 6 hours. Red dots signify genes that are candidates. Dots above the line indicate genes have a higher fitness on the mutant background, and dots below the line indicate they have a higher fitness on the wild type background.

Gene Validation

Synthetic dosage sick/lethal gene interactions for the S83L gyrA mutant

The 20 most promising genes, causing synthetic dosage sickness/ lethality for the mutant were determined, chosen based off their average epsilon value for trials 1 and 2 and having a binomial *P*-value less than 0.0001 in both trials (Table 4). A wide variety of genes are present on this list, which was unexpected, as I predicted that genes closely involved with *gyrA* and DNA supercoiling or repair such as *gyrB* would be high on the list.

Table 4: Synthetic dosage sick/lethal gene candidates on the mutant background, showing function of the gene and epsilon and p-values after 6 and 24 hours.

b number	Gene	Function	Average Epsilon- 6 hours	Average Epsilon- 24 hours
b0738	tolR	inner membrane protein, part of the Tol-Pal system	0.278043618	0.234819797
b2900	yqfB	conserved protein	0.223562067	0.087754309
b2545	yphC	No information	0.211246375	0.264819449
b3998	nfi	Endonuclease V, acts upon single stranded DNA and duplex DNA that has been damaged (including a hairpin or unpaired loop)	0.202736044	0.1916402
b3552	viaD	Sequence similarity suggests that it is a member of the OmpA-OmpF Porin Family	0.191057826	0.094662503
b3303	rpsE	Component of the 30S subunit of the ribosome	0.190052225	-0.00727411
b3503	arsG	Arsenate reductase	0.179781852	0.08978003
b0827	moeA	Molybdopterinmolybdenumtransferase mediates the ligation of molybdenum to molybdopterin	0.174150777	0.32932111
b3578	viaN	A membrane-spanning component of the YiaMNO Binding protein-dependent Secondary (TRAP) transporter	0.16436018	0.228274603

b1501	ydeP	Overproduction causes resistance to low pH	0.161955998	0.206295316
b4477	dgoA	2-oxo-3-deoxygalactonate 6-phosphate aldolase(KDPGal aldolase) catalyzes the final reaction in the degradation of D-galactonate, an aldol cleavage resulting in GA3p and pyruvate which enter central metabolism	0.156271134	0.039259243
b0151	fhuC	component of the iron (III) hydroxamate ABC transporter	0.155035436	0.095020778
b2074	mdtA	Component of the MdtABC-TolC multidrug efflux system	0.151622603	0.180502536
b0193	yaeF	No information	0.149621124	0.182415347
b1757	ynjE	Monomeric sulfurtransferase containing three rhodanese-like domains	0.14829441	0.127912771
b0188	tilS	tRNA ^{Ile} -lysine synthetase, the enzyme responsible for modifying the wobble base of the CAU anticodon of tRNA ^{Ile} at the keto group in position 2 of C34 such that it exhibits proper recognition of the AUA codon rather than the AUG codon.	0.146777066	0.004904406
b3459	panZ	involved in processing the PanD proenzyme	0.145382483	0.090776248
b0392	ykiA	No information	0.144177184	0.03328968
b2946	rsmE	ribosomal RNA small subunit methyltransferase E	0.143820367	0.195016559
b0087	mraY	phospho-N-acetylmuramoyl-pentapeptide transferase	0.143103899	-0.082112923

The top 5 genes were the following: *tolR*, with an average epsilon value of 0.2780, encodes an inner membrane protein part of the Tol-Pal system (Cascales, 2000). *yqfB*, with an average epsilon value of 0.2236, and *yphC* with an average epsilon value of 0.2112, both have unknown functions. *nfi*, with an average epsilon value of 0.2027, codes for Endonuclease V, which acts upon single stranded DNA and duplex DNA that has been damaged (Guo, 1998), and has a function that is more expected. *yiaD* with an average epsilon value of 0.1911, which from sequence similarity suggests it is a member

of the OmpA-OmpF porin family. Porins are implicated in quinolone resistance (Kishii, 2009)

The difference in fitness of each synthetic dosage lethal/sick gene candidate wild type background against the fitness on the mutant background after 6 hours and 24 hours was determined (Figure 8). The fitness of each gene is higher on the wild type background than it is on the mutant background.

Figure 8. Selection coefficients for 20 gene candidates that are synthetically dosage lethal/sick on the mutant S83L *gyrA* background at A) 6 hours and B) 24 hours. Fitness represented by adjusted mean selection coefficients for trial 1 and trial 2.

Beneficial gene interactions for the S83L gyrA mutant

The 20 most promising genes causing synthetic dosage sickness/ lethality for the wild type were determined (Table 5), indicating that they are compensatory or beneficial for the mutant, chosen based on their average epsilon value for trials 1 and 2 and having a binomial *P*-value in both trials less than 0.0001. I predicted that there would be compensatory mutations to aid in growth, as the S83L *gyrA* mutation creates defects in growth. The first gene showing compensatory epistatic interactions for the mutant is *hyfC* with an average epsilon value of -0.3038, which encodes for a hydrogenase, and aids the cell during stress (Andrews, 1997). *yniD*, with an average epsilon value of -0.2367, encodes for a predicted transmembrane segment (Hemm, 2008). *relB* is an interesting gene to fall in the top five, with an average epsilon value of -0.2341, is part of the RelE-RelB toxin-antitoxin system (Gotfredsen, 1998). The system is involved in the regulation of cell growth under conditions where nutrients are limiting. RelE stops RelE-mediated translation of RelB, which is a growth inhibitor (Gotfredsen, 1998). *udk*, with an average epsilon value of -0.2096, encodes a uridine-cytidine kinase which phosphorylates both uridine and cytidine (Neuhard, 1993). The last gene in the top five is *ybeF*, a secreted protein of an unknown function, with an average epsilon value of -0.2060.

Table 5: Compensatory or beneficial gene candidates on the mutant background, showing function of the gene and average epsilon for trials 1 and 2 after 6 and 24 hours. Binomial *P*-values between the mutant and wild type selection coefficients were all less than 0.0001

b number	Gene	Function	Average Epsilon- 6 hours	Average Epsilon- 24 hours
b2483	hyfC	Encodes a hydrogenase- helps deal with stress	-0.303831626	-0.119579935
b4535	yniD	Predicted transmembrane segment	-0.236687493	-0.07558418
b1564	relB	RelB is the antitoxin of the RelE- RelB toxin-antitoxin system. System is involved in the regulation of cell growth under conditions with limited nutrients. RelB stops RelE- mediated translation which inhibits cell growth.	-0.234090192	-0.043174849
b2066	udk	Uridine-cytidine kinase phosphorylates both uridine and cytidine	-0.209583761	-0.092617881
b1847	yebF	Secreted protein of unknown function	-0.205957346	-0.035158776
b1301	puuB	Probably the γ -glutamylputrescine oxidase in a putrescine utilization pathway	-0.205350162	-0.023333774
b1749	xthA	Exonuclease III, function to repair DNA sites	-0.202614673	-0.062639262
b1727	yniC	Sugar phosphatase- substrate is 2-deoxyglucose-6-phosphate	-0.201999657	-0.0197563
b1045	ymdB	Macrodomain family protein with O-acetyl-ADP-ribose deacetylase activity, expression occurs in early stationary phase and during cold shock	-0.196478814	0.052226489
b1760	ynjH	Unknown	-0.184262022	-0.052814044
b0508	hyi	Subunit composition of hydroxypyruvate isomerase	-0.181065434	-0.083609379
b1490	dosC	DosC protein is composed of an N-terminal sensory globin domain that binds heme in a apolar binding pocket, and a C-terminal	-0.180928265	-0.034473096

		GGDEF diguanylate cyclase domain. Over-expression of DosC leads to increased levels of c-di-GMP in the cell and has a deleterious effect on cell division		
b2665	ygaU	Induced under osmotic stress, thought to have a domain that binds to membranes, and another that is associated with bacterial cell wall degradation	-0.176639462	-0.031259026
b0939	elfD	Promotes biofilm formation	-0.167994951	-0.037402467
b4437	sibB	Regulatory RNA	-0.166724633	-0.042777505
b1090	plsX	fatty acid/phospholipid synthesis protein	-0.165512612	-0.030428212
b2201	ccmA	Protoheme IX ABC transporter-ATP binding subunit	-0.164982161	0.027872565
b4620	yjbT	Unknown	-0.164841862	-0.000411388
b1499	ydeO	DNA-binding transcriptional dual regulator, activates genes involved in the cellular response to acid resistance	-0.163457648	-0.063009177
b0685	ybfE	Levels of ybfE mRNA are increased by exposure to the DNA damaging agent mitomycin C, maybe involved in DNA repair	-0.160582876	-0.12238315

The difference in fitness of each gene candidate on the wild type background against the fitness on the mutant background after 6 and 24 hours was determined (Figure 9). The fitness of each gene is higher on the mutant background than on the wild type background.

Figure 9: Selection coefficients for 20 gene candidates that are beneficial on the mutant S83L *gyrA* background at A) 6 hours and B) 24 hours. Fitness represented by adjusted mean selection coefficients for trial 1 and trial 2.

Validation Experiments

In order to validate the high-throughput screen, growth curves were generated for single wild-type or mutant genotypes bearing over-expression constructs for all genes of interest. Table 6-8 gives the p-values determined from a two factor ANOVA test for maximum growth rate and optical density after 6 hours and maximum optical density after 24 hours for 16 gene candidates, when compared to the pseudogene *cybC*, on the mutant and wild type backgrounds. The first factor p-values reflect the effect of background genotype, and the second factor *P*-values reflect the effect of the over-expression plasmid.

Genes *nfi*, *yiaN* and *puuB* were determined to have significant epistatic interactions, from statistically significant differences between wild type and S83L *gyrA* mutant's maximum growth rate based on interaction p-values (Table 6). Genes *yaeF*, *dgoA*, *ydeP*, *ymdB*, *hyfC* and *ynjH*, were determined to have significant epistatic interactions, from statistically significant differences between wild type and S83L *gyrA* mutants OD after 6 hours, based on interaction p-values (Table 7). Genes *yaeF*, *nfi*, *fhuC*, *yiaN*, *ydeP*, *ymdB* and *hyfC* were determined to have significant epistatic effects, from statistically significant differences between the wild type and mutant OD after 24 hours, based on interaction *P*-values (Table 8).

Table 6. Two factor anova *P*-values to determine statistically significant differences in maximum growth rate to detect epistatic interactions. First factor *P*-value describes the effects of the overexpressed gene in comparison to the effects of an overexpressed pseudogene. Second factor *P*-value describes the effects of the background genotype.

Gene	Predicted Effect on Mutant Background	First Factor <i>P</i> -value	Second factor <i>P</i> -value	<i>Interaction P</i> -value
<i>moeA</i>	Synthetic dosage lethal/sick	0.00001	0.123	0.846
<i>yaeF</i>	Synthetic dosage lethal/sick	0.0512	0.472	0.319
<i>nfi</i>	Synthetic dosage lethal/sick	0.373	0.440	0.0221 ^a
<i>fhuC</i>	Synthetic dosage lethal/sick	0.0000177	0.00703	0.650
<i>rspE</i>	Synthetic dosage lethal/sick	0.444	0.155	0.181
<i>yiaN</i>	Synthetic dosage lethal/sick	0.00232	0.00000366	0.000000733 ^a
<i>yphC</i>	Synthetic dosage lethal/sick	0.297	0.654	0.298
<i>mdtA</i>	Synthetic dosage lethal/sick	0.0577	0.0606	0.798
<i>dgoA</i>	Synthetic dosage lethal/sick	0.142	0.00516	0.545
<i>yqfB</i>	Synthetic dosage lethal/sick	0.241	0.219	0.153
<i>ydeP</i>	Synthetic dosage lethal/sick	0.194	0.805	0.910
<i>ymdB</i>	Compensatory	0.613	0.218	0.0776
<i>puuB</i>	Compensatory	0.000595	0.000934	0.0355 ^a

<i>hyfC</i>	Compensatory	0.000301	0.0764	0.227
<i>udk</i>	Compensatory	0.465	0.0816	0.429
<i>ynjH</i>	Compensatory	0.170	0.0155	0.546

^a- signifies epistatic interaction detected

Table 7. Two factor anova *P*-values to determine statistically significant differences in optical density after 6 hours to detect epistatic interactions. First factor *P*-value describes the effects of the overexpressed gene in comparison to the effects of an overexpressed pseudogene. Second factor *P*-value describes the effects of the background genotype.

Gene	Predicted Effect on Mutant Background	First Factor <i>P</i> -value	Second factor <i>P</i> -value	Interaction <i>P</i> -value
<i>moeA</i>	Synthetic dosage lethal/sick	.000000000000356	0.989	0.438
<i>yaeF</i>	Synthetic dosage lethal/sick	0.00000000000127	0.171	0.0476 ^a
<i>nfi</i>	Synthetic dosage lethal/sick	0.000000875	0.0714	0.138
<i>fhuC</i>	Synthetic dosage lethal/sick	0.00000000994	0.0935	0.0518
<i>rspE</i>	Synthetic dosage lethal/sick	0.850	0.772	0.896
<i>yiaN</i>	Synthetic dosage lethal/sick	0.0000000000741	0.336	0.567
<i>yphC</i>	Synthetic dosage lethal/sick	0.110	0.418	0.387
<i>mdtA</i>	Synthetic dosage lethal/sick	0.00000000000419	0.330	0.138
<i>dgoA</i>	Synthetic dosage lethal/sick	0.0229	0.0767	0.0305 ^a

<i>yqfB</i>	Synthetic dosage lethal/sick	0.00000617	0.875	0.824
<i>ydeP</i>	Synthetic dosage lethal/sick	0.0000240	0.00000603	0.000004 ^a
<i>ymdB</i>	Compensatory	0.0000222	0.000373	0.000179 ^a
<i>puuB</i>	Compensatory	0.00689	0.236	0.181
<i>hyfC</i>	Compensatory	0.000000059	0.00451	0.00290 ^a
<i>udk</i>	Compensatory	0.00401	0.464	0.595
<i>ynjH</i>	Compensatory	0.000232	0.0188	0.0111 ^a

^a - signifies epistatic interaction detected

Table 8. Two factor anova p-values to determine statistically significant differences in optical density after 24 hours to detect epistatic interactions. First factor p-value describes the effects of the overexpressed gene in comparison to the effects of an overexpressed pseudogene. Second factor p-value describes the effects of the background genotype.

Gene	Effect on Mutant Background	Effect on Mutant Background	First Factor <i>P</i> -value	Second <i>factor P</i> -value
moeA	Synthetic dosage lethal/sick	0.000000000395	0.213	0.365
yaeF	Synthetic dosage lethal/sick	0.0000685	0.0067	0.0219 ^a
nfi	Synthetic dosage lethal/sick	0.0000000532	0.0000789	0.000188 ^a
fhuC	Synthetic dosage lethal/sick	0.00001	0.00447	0.00646 ^a
rspE	Synthetic dosage lethal/sick	0.652	0.373	0.825
yiaN	Synthetic dosage lethal/sick	0.0000000927	0.0000000434	0.000000036 ^a
	Synthetic dosage lethal/sick	0.0153	0.0351	0.0903

yphC				
mdtA	Synthetic dosage lethal/sick	0.314	0.866	0.533
dgoA	Synthetic dosage lethal/sick	0.00305	0.117	0.34
yqfB	Synthetic dosage lethal/sick	0.000244	0.913	0.523
ydeP	Synthetic dosage lethal/sick	0.00000323	0.00000895	0.0000196 ^a
ymdB	Compensatory	0.0523	0.000334	0.000844 ^a
puuB	Compensatory	0.372	0.236	0.0622
hyfC	Compensatory	0.00000222	0.000735	0.00292 ^a
udk	Compensatory	0.372	0.236	0.0622
ynjH	Compensatory	0.0319	0.0639	0.163
a- signifies epistatic interaction detected				

Not all the genes behaved as expected. The predicted epistatic effect to what we actually observed were compared (Table 9). Only two synthetic dosage lethal/sick interactions were confirmed, for *yiaN* and *nfi*. The rest of the genes seemed to grow equally well or better on the mutant background, when compared to the wild type background. Four out of the five compensatory genes showed compensatory effects.

Table 9. Predicted effect of over-expression of each gene on mutant background, description of the actual effect on both the mutant and wild type background, and confirmation of epistatic interaction

Gene	Predicted Effect on Mutant Background		Validation
	Based on Screen	Effect seen during pure culture growth	
moeA	Synthetic dosage lethal/sick	Greatly decreased growth on both the mutant and wild type background. Slightly higher OD after 24 hours on the mutant background.	Not Confirmed
yaeF	Synthetic dosage lethal/sick	Slightly impaired growth on the wild type background	Not Confirmed
nfi	Synthetic dosage lethal/sick	Slower growth on the mutant background, but lower optical density on the wild type background after 24 hours	Confirmed
yiaN	Synthetic dosage lethal/sick	Lethal on the mutant background, unimpaired growth on the wild type background	Confirmed
dgoA	Synthetic dosage lethal/sick	Slightly impaired growth on the wild type background, only after 6 hours	Not Confirmed
ydeP	Synthetic dosage lethal/sick	Impaired growth on the wild type background	Not Confirmed
ymdB	Compensatory	Increased growth on the mutant background at 6 and 24 hours	Confirmed
puuB	Compensatory	Higher maximum growth rate on mutant background, but Optical Density evens out after 24 hours	Confirmed
hyfC	Compensatory	When compared to pseudogene, both the wild type and mutant backgrounds experienced a decrease in growth, but on the wild type the decrease was much greater	Confirmed
ynjH	Compensatory	Impaired growth on wild type background at 6 hours, but evens out after 24	Confirmed

The average growth curves for the pseudogene *cybC* and candidate genes *yiaN* and *nfi* were plotted (Figure 10). Overexpression of *yiaN* is lethal on the mutant background, but not on the wild type background, confirming the epistatic interaction detected from the screen.

Figure 10. Optical density growth curves of liquid culture for 24 hours at 37 °C of the 1 confirmed genes *yiaN* and *nfi*, that shows lethal effects on the mutant background, and one pseudogene, *cybC*. Over-expression plasmids carrying these genes have been transformed onto both the wild type and mutant backgrounds.

The average growth curves for pseudogene *cybC* and four compensatory genes *puuB*, *ynjH*, *hyfC* and *ymdB* were plotted (Figure 11). All four genes show compensatory

effects for the mutant, confirming the epistatic interaction detected from the screen.

hyfC, while showing an impaired growth curve for both the mutant and wild type when compared to the growth curve of *cybC*, is more detrimental for the wild type than the mutant background.

Figure 11. Optical density growth curves of liquid culture for 24 hours at 37 °C of the 4 confirmed genes *puuB*, *ynjH*, *hyfC* and *ymdB* that show compensatory effects on the mutant background, and one pseudogene, *cybC*. Over-expression plasmids carrying these genes have been transformed onto both the wild type and mutant backgrounds.

Based on the results from the validation experiments, the over-expression of *yiaN* gives the most striking lethal effect, almost completely preventing growth when expressed on the mutant background, but having no effect on the wild type background. Candidates picked from the screen do not always work, at least not for this type of validation assay. The screen may favor detecting compensatory epistatic interactions.

Discussion

I have created a high-throughput screen using next generation sequencing to study epistatic interactions in *E. coli*. I used changes in plasmids carrying over-expression genes in a population of wild type and mutant cells to infer epistasis. This was determined by sequencing at different time points, to establish selection coefficients to be able to calculate epistasis for each over-expression plasmid covering 4123 different ORFs in *E. coli* on two different backgrounds. We compared over-expression plasmid frequency present on the wild type MG1655 *E. coli* background to a mutant *E. coli* background containing a of S83L *gyrA* point mutation, which provides antibiotic resistance to ciprofloxacin. This screen detected 216 epistatic interactions, both synthetic dosage lethal/sick or beneficial for the mutant strain.

We validated the 6 epistatic interactions determine from this screen using pure culture 24 hour growth curve assays, measuring maximum growth rate, and optical density after 6 and 24 hours. We validated two synthetic dosage lethal/sick genes, *nfi* and *yiaN* and four compensatory genes out of 10 in total tested. This is very promising in validating the results of our screen. The 4 that were not confirmed does not discredit the 216 interactions detected by our screen, and may indicate the need for competition in our validation assays.

High Throughput Method

Comparison to SGA

Studying epistatic interactions in the past has been tedious and time consuming. I set out to create a high-throughput screen with the aims of being fast, easy to perform, and quantitative, to analyze thousands of gene interactions at once.

Synthetic genetic arrays (SGA) are the current standard method to study gene interactions. These screens often look for synthetic dosage lethal/sick genetic interactions and can be performed in both *E. coli* and *Saccharomyces cerevisiae*. In yeast, SGA is typically carried out using an automated robotic system and knockout libraries, mated together to create haploid double mutants (Costanzo, 2011) (Typas, 2008). In *E. coli*, double mutants are created by conjugation between query strains and libraries of single gene knockouts. In both cases, epistatic interactions are inferred by either using visual inspection looking at colony size evaluated by a computer program, or growth in liquid media and measuring optical density as a measure of fitness. In the past, the lack of fully annotated genomes for bacteria has been a barrier, but now the availability of fully annotated genomes for many bacterial species including *E. coli*, makes conducting this screen possible.

The screen described here is completely different from the synthetic genetic arrays currently used, and takes advantage of next generation sequencing. It has been designed for high-throughput use in bacteria, but could be applied to yeast or other cellular systems. Our screen makes use of inducing the ASKA over-expression plasmid library on an antibiotic resistant (mutant) background, and comparing the survival of over-expression plasmids on the mutant background versus the wild type background.

Fitness of different over-expression clones is detected using high throughput sequencing, and observing changes in frequency.

This is different from SGA methods in several ways. Creating a pool of all the over-expression plasmids from the ASKA library and transforming them onto the mutant and wild type backgrounds relatively easy. The plasmid pool can be generated quickly by combining bacteria from a 96 well plate, where each well contains a unique over-expression plasmid, and extracting the plasmids all at once. If using the ASKA library, this only requires approximately 60 plasmid extractions. A single transformation can be performed with the pooled plasmids, creating ~4123 double mutants simultaneously.

I was conscious of the fact that I could be creating triple or quadruple mutants, which would skew my results. To ensure this was not the case, I selected and picked colonies after a transformation, and from my random selection of approximately 30 colonies, only single plasmid transformations were detected based on plasmid extractions, digesting the plasmids and using gel electrophoresis to determine the insert size.

Once the double mutants are made, the next difference from traditional SGA screens is how they are grown. SGA screens grow each double mutant individually, often as colonies on agar plates. By contrast, here the double mutants consisting of the over-expression plasmids on the mutant and wild type backgrounds, are grown and induced together, in one flask in liquid culture. The use of an inducible plasmid allows synthetic dosage lethal/sick phenotypes to be detected. The phenotype as a result of over-expression of the plasmids on each background would not occur until induced in

the experiment, which is not possible using a double knockout screen, where if a lethal interaction occurs, results immediately in a lack of growth. The ability to induce expression lets us control when synthetic dosage lethal interactions can be seen.

Instead of having to plate out every combination of the 4123 over-expression plasmids on each background, and measure epistasis through colony formation, these double mutants were grown together, and culture samples were taken at various time points during the experiment, including before induction. Plasmids were then extracted from these samples and sequenced using an Illumina Miseq. Sequence data were then analyzed to determine the frequency of each over-expression clone over time. This is another key difference between traditional SGA and my high throughput method. In other SGA screens, interaction is detected by differences in colony size, while in my screen interaction, is detected using quantitative measures in the change of plasmid frequency over time. Plasmid samples before induction gave me a baseline for what over-expression plasmids on each background were present before over-expression was induced. By taking samples at various time points, I can get a better idea of what is happening during log growth and stationary growth, something that is not offered by traditional SGA screens, but may give important insight as epistatic interactions may occur at different stages in bacterial growth.

Finally, and maybe the most important difference between this screen and traditional SGA screens, is the ability to quantify epistatic interactions in a way that does not rely on visual analysis or optical density. Instead, using a bioinformatics workflow the gene and the amount present in each sample can be determined from the

sequencing data, and change in frequency over time can be detected. Epistatic interactions are then inferred by differences in selection coefficient between the mutant and wild type backgrounds.

The amount of time it takes to perform this experiment is minimal once the necessary requirements are obtained, particularly the plasmid pool and query mutants. The actual experiment takes a day to perform, starting with the transformation of the plasmid library. Samples can be collected as frequently as desired, with only the pre-induction and an end time point being necessary. I believe it is worthwhile to include samples and therefore plasmid collection at multiple time points, including 6 hours. Most of the compensatory epistatic interactions were detected during log phase from the sequencing results at 6 hours, but not after 24 hours. Additionally, by sequencing additional time points, a more accurate selection coefficient may be determined, as more time points provide a clearer understanding of how frequency of the over-expression plasmids is changing over time

Illumina sequencing has a turn around time as quick as three days, including sample preparation. The bioinformatics workflow may be automated, making hands on data analysis minimal. Therefore, the screen makes it very quick and easy to query mutants of interest, testing ~4123 gene interactions at once.

The cost of this experiment varies with the complexity of the experiment. The ASKA library is reasonable, costing only approximately \$500 to obtain. The real expense is the next generation sequencing platform, with Illumina Miseq currently selling for about \$100,000. The materials to perform a Miseq run costs approximately \$2000. We

ran 9 samples per run, using two runs to complete this project. The robotic equipment necessary to run SGA fully automated is similar to the cost of an Illumina Miseq. Additional software and hardware, needed to evaluate colony size, represent added costs. Both with high start up costs, Illumina Miseq's have the benefit of being multifunctional, which may justify the cost, but have higher repetitive costs associated with running the actual experiment.

While synthetic lethal/synthetic dosage lethal or very apparent synthetic sick interactions can be detected with SGA, subtle epistatic relationships may be detected more easily with the current screen. Quantitatively, we were able to detect a range of epistatic interactions, with 216 both synthetic dosage lethal/sick and beneficial interactions detected for the mutant. Most gene interactions were subtle, only being slightly beneficial or slightly detrimental, with few being extreme. This ability to detect a range of epistatic interactions may give a more complete picture of the distribution of epistatic effects. Additionally, the added aspect of competition between the 4123 different over-expression plasmids on the mutant and wild type backgrounds may provide an environment where slightly detrimental combinations resulting in negative epistatic interactions are selected against, and are out-competed by neutral or beneficial combinations resulting in epistatic interactions. These subtle interactions may not be easily detected otherwise.

Challenges and Issues

There are challenges and issues that still need to be worked out, while we develop this screen further. There was a lack of consistency between the three trials performed in this experiment. Trials 1 and 2 were very consistent, as seen in figure 5, but trial 3 produced different results. We normalized the selection coefficients for each gene by subtracting the median pseudogene selection coefficient, calculated from 4 putative pseudogenes. We found there was a trend for most combinations of the over-expression plasmids on the mutant and wild type backgrounds to die out by 24 hours, even if over-expression gene was neutral, as more beneficial gene combinations were selected for. This may not be the best way to normalize for the competition that occurs, and overexpressing the pseudogenes, even though they have no predicted effect may have varying degrees of stress on the cell, such as costs associated with increased transcription.

Another issue is that currently we have to sequence the plasmid vector to be able to sequence the gene inserts. This means we are removing up to 90% of our sequence data in our bioinformatics workflow after sequencing. A potential way of cutting down costs and being able to sequence more samples at once is to remove the vector before sequencing. One way to do so is through biotinylation. By creating a biotinylated vector without an insert, and mixing it with digested plasmid pools, which have been cut between the gene insert and the rest of the vector, the biotinylated vector may stick to the vector backbone. Streptavidin beads can then be used to pull out the hybridized vector, only eluting the insert. With the vector removed prior to

sequencing, potentially 9 times more samples could be assessed during one sequencing run, greatly reducing costs.

Finally, the biggest challenge is: How well does it work? Results in preliminary gene validation experiments are encouraging, even though validation using 24 hour growth curves produced mixed results. I tested five genes that were detected as beneficial on the S83L background in the high-throughput screen, and growth curve data confirmed the findings (Figures 10 and 11). The synthetic dosage lethal/sick epistatic interactions on the mutant background were less clear-cut, with only one gene, *yiaN*, clearly showing a synthetic dosage lethal interaction on the mutant but not on the wild type background. Additionally, *nfi* showed decreases in maximum growth rate on the mutant background when compared to the wild type background, and although having similar maximum optical density, maximum growth rate may be more descriptive than optical density in describing defects in growth in a pure culture.

The validation experiment, where culture is pure, and only transformed with the over-expression plasmid of interest provides a very different environment than the high-throughput screen. The experimental environment of having all 4123 over-expression plasmids on either the wild type and mutant backgrounds provides competition between clones. This competition may then exacerbate only subtle differences in growth, resulting in slightly detrimental gene interactions being out competed. Using a pure culture in rich medium may make subtle growth differences caused by epistatic effects of the overexpressed gene hard to distinguish between the wild type and mutant strains, because of the lack of competition. Differences in growth rate are probably

more important in terms of selection, in a competitive environment. Therefore using differences in maximum growth rate to confirm epistatic interactions may be more accurate than just looking at optical density.

Additional Uses- Evolution Experiments

Although this experiment was designed as a high-throughput method to discover epistatic interactions, since different gene combinations are competing in the same environment, additional evolutionary questions may be answered using this screen. Many important questions about adaptation remain unanswered, including whether the way a population adapts over time changes as it approaches a fitness optimum (Orr, 2005). This question could be answered with the current screening strategy using multiple strains with different relative fitnesses. We could determine the distribution of selection coefficients for each strain, with the prediction that low fitness strains should have more available beneficial mutations, and mutations with larger effect sizes.

Additional Uses- Antibiotic Drug Targets

An important application of this screen is the potential to find candidate genes that, when over-expressed, are deleterious on an antibiotic resistant background, but not on a wild type background. Drugs could be then designed to mimic the effect of the overexpressed gene, and create very specific drugs that when used in combination with

older antibiotics, may be able to fully eliminate both antibiotic and non-antibiotic resistant strains of pathogenic bacteria.

Additional Uses- Cancer Drug Discovery

This screen is not limited to antibiotic drug discovery in bacteria. There is potential to use epistatic interactions to discover drug therapies against cancer cells. One well/one-gene screening techniques have been used to identify synthetic lethal gene interactions in cancer cells. For example, Whitehurst et al. (2007) identified gene targets that reduced cell viability in less than lethal concentrations of paclitaxel, a drug that is regularly used in the treatment of non-small-cell lung cancer (Whitehurst, 2007). They used a library of siRNAs to target 21,127 human genes against NCI-H1 155, a human non-small-cell lung cancer line. Using pools of 4 different synthetic siRNA duplexes against a single gene target (to ensure knockdown), they measured changes in cell viability by monitoring ATP production (Whitehurst, 2007). From this they could identify genes that modified paclitaxel sensitivity, and potentially create drug targets to mimic the knockdown action of the screen.

In applying our screen, cancer cells could be transfected by a pooled siRNA library, grown over a set period of time and RNA extracted, siRNA detected using PCR and specific primers and then sequenced. Looking for changes in abundance of siRNA, epistatic interactions could be inferred. During the growth period, different drugs such as paclitaxel or experimental conditions could be applied. Essentially similar to Whitehurst's screen in expected results, applying our screen may speed up detection, as

21,217 interactions could be analyzed at once. Additionally, our screen could be applied to yeast with classic cancer mutations and anti-cancer drugs.

Additional Uses- Orphan Disease Treatment discovery

A unique and unusual application of this screen may be in finding treatment options for orphan diseases. Not only has the discovery of new antibiotics slowed, the discovery of drugs to correspond with new diseases has also slowed (Brinkman, 2006). Monogenic disorders, where a mutation in one gene results in disease, are often rare and understudied, because of the small population that they affect (Brinkman, 2006). Once mutations leading to disease are identified, the next step is to understand the function of the gene and the pathway that it is part of (Brinkman, 2006). Describing the epistatic interactions involved in that pathway may be key to finding drugs to reverse the phenotype of the disease.

Guernsey et al. (2009) showed that experiments done in yeast could be used to understand and develop treatment for the orphan diseases in humans (Guernsey, 2009). They studied a rare form of sideroblastic anemia caused by mutation in the gene encoding the protein SLC25A38, which is an erythroid specific mitochondrial carrier family protein (Guernsey, 2009). Using yeast, it was determined that this mutation affects glycine transport into the mitochondria, and therefore treatments could be designed to reverse this defect (Guernsey, 2009). Since these disorders are monogenic, if present in a highly conserved gene that is orthologous in bacteria or yeast, epistatic screens could determine the mode of action of disorders and then work towards

reversing them. By producing mutant bacterial or yeast strains that mimic the genetic disorder, we could tease out epistatic interactions and potentially find compensatory genes that, when overexpressed, reduce the fitness defects of such disorders. These compensatory genes could then become drug targets.

Many genetic questions could be asked and answered with this high-throughput screen. Whether based on evolutionary, genetic, or medical questions, due to the extensive epistatic nature of genetic systems, understanding how genes interact provides many questions and potential answers in numerous fields in biology. The quick turnover time, relative ease, and large amount of data produced with each experiment makes this screen powerful and will complement traditional SGA methods and the data already determined from traditional screens. The application to other organisms such as other bacterial, yeast or even cancer cell lines indicates the broad ways this screen may be applied.

Discovery of Gene Candidates for Drug Targets and Beneficial Mutations

Antibiotic Resistance- A Growing Problem

It is well established that antibiotic resistance is a growing problem, with the World Health Organization warning that we are moving towards a “post-antibiotic era” (World Health Organization, 2014). Despite the great need for new antibiotics to be discovered, tested and produced, pharmaceutical companies have not been moved into action. This is in part due to the inherent risk of resistance involved in producing drugs against microorganisms that can evolve. Additionally, developing antibiotics realizes a

lower profit, compared to other types of medication, and the strictness in new drug approval from the FDA, makes putting new antibiotics onto the market more expensive than in the past.

Antibiotic Gene Candidates detected by Screen

The S83L *gyrA* point mutation provides antibiotic resistance against fluoroquinolones, and specifically ciprofloxacin. We have detected up to 149 synthetic dosage lethal/sick epistatic interactions in double mutants containing an over-expression plasmid with the candidate gene on the S83L *gyrA* antibiotic resistant background. Genes interacting with the S83L mutation have a variety of functions, including involvement with DNA (*nfi*, *yphC*), porin formation (*yiaD*), carbon source usage (*yiaN*), and metal acquisition (*dgoA*). Most of these interactions were not expected, as we expected to find genes involved in supercoiling (e.g. *gyrB*) or in DNA repair.

yiaN

yiaN is the main potential drug target identified in this screen. *yiaN* had nearly neutral selection coefficients of -0.013 and -0.016 on the wild type background after 6 and 24 hours respectively, and negative selection coefficients of -0.177 and -0.244 on the mutant background after 6 and 24 hours respectively. The epsilon values measuring epistasis were 0.164 after 6 hours and 0.228 after 24 hours. The results from the validation experiments using 24-hour growth curves show that overexpressing *yiaN* on the mutant background is lethal, while showing no detrimental growth effects on the

wild type background (Figure 10). *yiaN* is therefore a strong candidate for targeting the S83L *gyrA* mutant.

yiaN encodes for large membrane-spanning protein and is part of the YiaMNO tri-partite ATP-independent periplasmic (TRAP) transporter. YiaN is the small membrane-spanning protein, and YiaO is the periplasmic solute-binding protein (Plantinga, 2004). The TRAP transporter works without using ATP, transporting the solute across the membrane by taking advantage of thermodynamically favorable transport of ions such as Na⁺ and H⁺. Due to dependence on these counter ions, TRAP transporters can work bidirectionally if necessary (Mulligan, 2010).

The YiaMNO transporter binds to 2,3-Diketo-L-gulonate as its substrate and transports it across the membrane (Mulligan, 2010). 2,3-Diketo-L-gulonate is formed when ascorbic acid is solubilized, and subsequently oxidized in water and can be used as a carbon source by *E. coli*. Additionally, transcription of the operon increased mildly in the presence of L-ascorbate (Ibañez E, 200) but L-ascorbate is not thought to be carried through the membrane by this transporter (Campos, 2007).

Deletion of the YiaMNO genes prevents *E. coli* from transitioning into stationary phase, reduces salt tolerance, and reduces its ability to form biofilms, indicating the transporter may do more than just transport 2,3-Diketon-L-gulonate (Plantinga, 2005). . YiaN itself is not known to regulate transcription of this operon, and presumably of any operon, and is not known to bind to any additional molecules in the cell.

It is not immediately obvious how over-expressing *yiaN* is creating a lethal interaction with the S83L *gyrA* mutant. One possibility is that an increase in YiaN offsets

the stoichiometry of the pump, knocking out its function. Alternatively, a potential in the increased import of 2,3-Diketo-L-gulonate or L-ascorbate due to more YiaMNO transporters being included in the cell membrane, may be increasing stress on the cell, by increasing oxidative stress from the transported substrate (Campos, 2007).

nfi

nfi is the second potential drug target identified by this screen. Seen in figure 10, in pure culture validation experiments, *nfi* over-expression did not produce the same level of synthetic dosage lethal interaction seen by *yiaN*, and eventually reached a greater OD than the wild type, but did result in a statistically significant decrease in maximum growth rate.

In the high-throughput screen, the over-expression of *nfi* had nearly neutral selection coefficients on the wild type background of -0.048 after 6 hours and -0.022 after 24 hours. On the mutant background it had negative selection coefficients of -0.250 and -0.214 after 6 hours and 24 hours respectively. The epsilon value, indicating epistasis, was 0.203 after 6 hours and 0.192 after 24 hours.

nfi encodes Endonuclease V (EndoV), which repairs damaged DNA from a variety of damaging agents such as UV and low pH, and especially by nitrous acid (Guo, 1998) by recognizing deoxyinosine on double and single stranded DNA (Yao, 1996). Over-expression of *nfi* on wild type backgrounds was not found to be lethal (Guo, 1998). It does not remove the damaged bases, but it is believed to recruit another protein to do so (Yao M, 1996).

Since the GyrA-GyrB complex is involved with the supercoiling of the DNA, it must cut the double strand to do so. The S83L mutation is in the active site, where the DNA binds, and potentially interferes with the GyrA subunit. Prolonged cut DNA inside the cell, combined with the overproduction of EndoV, may interfere with the GyrA-GyrB complex to proceed with supercoiling and therefore cell replication.

S83L-specific beneficial mutations

Often, but not always, antibiotic resistance mutations may have fitness costs on the cell. Because of these fitness costs, it is interesting and beneficial to understand which overexpressed genes give beneficial epistatic interactions on the mutant background. This may provide information in to how antibiotic resistance negatively effects fitness, as these mutations are being selected for as they compensate against such costs, or give us information on the type of secondary mutations that may arise when groups of antibiotic resistant bacteria compete against each other. For the S83L *gyrA* mutation in particular, there is mixed evidence that this mutation has costs associated with it. Using competition assays (data not shown) performed by Kimberley Seguin, the S83L *gyrA* mutant had the same fitness as the wild type strain when grown in liquid LB. Some published research supports this, saying that there is no fitness cost for S83L mutations (Marcusson, 2009). Although they also say there are for mutations in *marR*, another common gene to acquire point mutations providing antibiotic resistance (Marcusson, 2009). Alternatively, other studies have said there is fitness costs associated with the S83L *gyrA* mutation (Andersson, 2010).

I carried out validation experiments for five genes that showed benefits specific to the S83L mutant. Four out of these five genes showed such benefits in pure culture 24 hour growth curves. They were *puuB*, *ymdB*, *hyfC* and *ynjH*. *puuB* was validated based on maximum growth rate, while the others were validated by having a greater optical density when compared to the wild type expressing the gene at either 6 or 24 hours.

puuB

The over-expression of *puuB* on the wild type background resulted in a negative selection coefficient after 6 hours of -0.121 and evened out to be neutral after 24 hours at -0.039. On the mutant background the selection coefficients remained neutral at 0.0846 and -0.016 after 6 and 24 hours respectively. The epsilon values were -0.205 and -0.0233 after 6 and 24 hours respectively. This indicated that epistatic interactions were experienced after 6 hours, but not after 24 hours. This may indicate that these gene interactions were important in the log phase. Consistent with this hypothesis, maximum growth rate was significantly greater when *puuB* was expressed on the mutant background than on the wild-type background.

puuB is expected to be a gamma-glutamylputrescine oxidase (Kurihara, 2004). Putrescine is a polyamine that is often found when cells are proliferating or under stressful conditions in plants and animals and is normally found in *E. coli* during replication and growth (Kurihara, 2004). Since *puuB* is important for cell growth and can utilize putrescine during nutrient starvation, if the mutant is experiencing growth

defects due to the S83L *gyrA* mutation, increased gamma-glutamylputrescine oxidation may be aiding in some way to compensate for this.

ymdB

The over-expression of *ymdB* on the wild type background resulted in neutral selection coefficients at both 6 and 24 hours of -0.004 and -0.009 respectively and on the mutant background, 0.191 at 6 hours and -0.061 after 24 hours, resulting in a positive selection coefficient during log growth, evening out to a neutral selection coefficient after 24 hours. The epsilon values were -0.196 and 0.052 after 6 and 24 hours respectively, indicating like *ymdB* epistatic interactions were experienced during log phase.

ymdB inhibits the enzymatic activity of RNase III by interacting with it directly. It also has *O*-acetyl-ADP-ribose deacetylase activity and over-expression of *ymdB* inhibits biofilm formation (Kim, 2013). Kim reports that *ymdB* strongly regulated 129 genes indicating that *ymdB* is involved in many cellular processes, although they did not detect an interaction with *gyrA* (Kim, 2013). Since *ymdB* is extensively modulates many pathways, including biofilm formation, this may be why over-expression is beneficial for the mutant.

ynjH

The over-expression of *ynjH* on the wild type background resulted in neutral selection coefficients at both 6 and 24 hours of -0.076 and 0.001 respectively, and on the mutant background, 0.108 and 0.054. The epsilon values were -0.184 and -0.052, showing like *ymdB* and *puuB*, epistatic interactions were exacerbated during log phase.

ynjH is a predicted protein, dependent on RNA polymerase sigma 28 for transcription (Yu, 2006). It is conserved through many species of bacteria, but the function is unknown, so how expression is causing a beneficial gene interaction on the mutant background remains undetermined.

hyfC

The overexpression of *hyfC* on the wild type background resulted in having a negative selection coefficient on the wild type background at 6 hours of -0.234 and a neutral coefficient after 24 hours of -0.039. On the mutant background after 6 and 24 hours the selection coefficients were nearly neutral being 0.070 and 0.081 respectively. It had an epsilon value of -0.304 and -0.120 after 6 and 24 hours respectively.

The over-expression of *hyfC*, and the detection of this epistatic interaction, is most likely not a true beneficial interaction on the mutant background, and rather just a negative interaction on the wild type background. When looking at Figure 11, it is evident that the mutant with *hyfC* expression had a decrease in growth, but the decrease was not as great in the mutant, making over-expression on the mutant background more neutral or less deleterious than it is on the wild type background.

hyfC is part of the *hyf* operon, which codes for 12 genes that make up hydrogenase 4, a 9 subunit protein that interacts with formate dehydrogenase (Andrews, 1997). This interaction creates an active formate hydrogenlyase complex that converts formate into dihydrogen and carbon dioxide (Andrews, 1997). In *E. coli* the *hyf* operon is probably silent, as it has not been found to be expressed at significant levels, and expression required growth at a pH of 7.5 and for cells to be stressed hyperosmotically (Self, 2004). Since it may not be normally expressed, and it is part of a nine gene complex, it is not apparent why the expression of *hyfC* alone is producing epistatic interactions in this screen.

Other potential drug targets

Only a handful of potential gene candidates for drug targets were tested initially. Positive validation outcomes from only a few synthetic dosage lethal/sick interactions, but many compensatory interactions, may lead us to evaluate the type of tests we use to validate these results. Synthetic dosage lethal/sick interactions may only be apparent in environments with competition present.

Comparison to current methods in antibiotic target discovery

Currently there are many methods employed to combat resistance found in pathogenic bacteria. Scientists are trying to revive old drugs by either modifying them by changing their structure, combining them with new therapies such as peptides that encourage the activity of the immune system, to increase activity against pathogens,

removing the toxins bacteria make by creating nanosponges, and moving away from broad spectrum antibiotics to very specific tailored antibacterial proteins (Gammon, 2014).

Not only does our screen become an important stand-alone tool in the search for new antibiotics, combined with some of these new therapies and detection methods, it becomes even more powerful. Running this screen using antibiotic resistant mutants in environments with antibiotics, may allow us to successfully detect epistatic interactions that are lethal and reverse the effects of antibiotic resistance mutations. Running this screen with other drug combinations or treatments listed above may allow us to further design drugs to exacerbate their effects, or to improve these new treatments.

The potential of this screen is that drugs may be one day be tailored to the needs of the patient. Based on resistance mutations in the bacteria present, therapy may be designed to target only the disease causing bacteria, with the aim of preventing resistance as much as possible.

Impact on Evolutionary Theory

As stated earlier, traditionally epistasis has not been considered when producing mathematical models to describe evolutionary processes. The general view is that evolution occurs based stepwise substitutions, where genes have the same effect on every background. By contrast, if epistasis is common, a mutation's effects are potentially dependent upon many genes interacting as a system (Hansen, 2013). In

terms of selection, population genetics generally assumes that epistasis only shapes populations in the short term (Hansen, 2013). If epistasis is widespread, a gene-by-gene model may be inaccurate in describing how selection occurs, and instead models should assume that the effects of mutations are dependent on the genetic background in which they occur (Hansen, 2013).

This screen described 216 potential epistatic interactions for one point mutation in a highly conserved gene. There was a large range in strength of the epistatic interactions, showing that the antibiotic point mutation interacts with the over-expression plasmid products, with a range of effects on fitness. Most gene interactions were subtle, only being slightly positive or slightly negative, with few being extreme in their phenotype. With 4123 genes being tested on the mutant and wild type background, 5% showed evidence for epistasis, which is much higher than the median 22 predicted by Costanza (Costanza, 2010). An interesting difference of our experiment that is not typically seen in large SGA experiments is the type of mutations we chose to investigate. Our background mutation was a point mutation in an essential gene, rather than the deletions in non-essential genes used in SGA screens. Choosing a point mutation, it was expected that only a few genes, mainly tied to DNA replication and repair, would show epistatic interactions, but this was not the case. I believe that this shows that epistatic interactions are elaborate, often not easily predicted and extensive.

My results, while providing the ability to detect more subtle epistatic interactions, largely correspond with results from other studies. Large SGA screens in yeast have produced varying results, but with the underlying theme that most genes do

display epistatic interactions with other genes, with one study citing the median value of 22 interactions per gene (Costanza, 2010). Small scale experimental evolution experiments as early as the 1980's, have shown the importance in epistatic interactions in providing compensatory mutations in populations experiencing fitness trade-offs due to mutations that provide resistance or some other benefit, but reduce growth or competitive fitness (Lenski, 1988). But large-scale studies focused on defining the extent of epistasis and effect on adaptation are missing in the literature, despite the existence of epistasis being readily accepted and mattering as a key component in genetic systems.

The biggest impact that understanding the distribution of epistatic effects may have on evolution, is the ability to describe how previous selected mutations shape future selection, and therefore the ability to determine or predict the molecular trajectory of a population. These adaptive walks describe how mutations are selected in a stepwise fashion, one after another (Martinez, 2014). It is unknown to what extent previous mutations shape succeeding evolutionary trajectories, but the current finding of extensive epistasis suggests that these effects could be substantial

.

Conclusion

Epistasis is increasingly recognized as important in evolutionary theory, systems biology, and medicine. As such, clear and straightforward high-throughput screens are necessary to characterize the distribution of epistatic effects. We have successfully developed a high-throughput screen to do so, making use of next generation

sequencing, resulting in a screen that is fast, performed with ease, and with adjustments could become relatively low cost.

By using a clinically relevant point mutation (*gyrA* S83L) in *E.coli*, resulting in quinolone resistance, we were able to use this screen to find candidate drug targets. Over-expression of *yiaN* and *nfi* was deleterious only on the mutant background. We were also able to identify mutations that were beneficial on the mutant background, indicating possible secondary mutations that may arise in these antibiotic resistant strains.

The distribution of epistatic effects was determined to consist of 216 genes for this point mutation. This makes up 5% of the genes tested, and indicates that epistasis is an important part of genetic systems, and therefore should be considered in evolutionary modeling, where it is often ignored.

References

- Andersson, D. I., & Hughes, D. (2010). Antibiotic resistance and its cost: is it possible to reverse resistance?. *Nature Reviews Microbiology*, 8(4), 260-271.
- Andrews, S. C., Berks, B. C., McClay, J., Ambler, A., Quail, M. A., Golby, P., & Guest, J. R. (1997). A 12-cistron Escherichia coli operon (hyf) encoding a putative proton-translocating formate hydrogenlyase system. *Microbiology*, 143(11), 3633-3647.
- Butland, G., Babu, M., Díaz-Mejía, J. J., Bohdana, F., Phanse, S., Gold, B., ... & Emili, A. (2008). eSGA: E. coli synthetic genetic array analysis. *Nature methods*, 5(9), 789-795.
- Bell, G. (1997). *Selection: the mechanism of evolution*. Springer.
- Bender, A. L. A. N., & Pringle, J. R. (1991). Use of a screen for synthetic lethal and multicopy suppressor mutants to identify two new genes involved in morphogenesis in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 11(3), 1295-1305.

- Brinkman, R. R., Dubé, M. P., Rouleau, G. A., Orr, A. C., & Samuels, M. E. (2006). Human monogenic disorders—a source of novel drug targets. *Nature Reviews Genetics*, 7(4), 249-260.
- Campos, E., Montella, C., Garces, F., Baldoma, L., Aguilar, J., & Badia, J. (2007). Aerobic L-ascorbate metabolism and associated oxidative stress in *Escherichia coli*. *Microbiology*, 153(10), 3399-3408.
- Cascales, E., Gavioli, M., Sturgis, J. N., & Lloubes, R. (2000). Proton motive force drives the interaction of the inner membrane TolA and outer membrane pal proteins in *Escherichia coli*. *Molecular microbiology*, 38(4), 904-915.
- Clardy, J., Fischbach, M., & Currie, C. (2009). The natural history of antibiotics. *Current biology: CB*, 19(11), R437.
- Clark, A. G., & Wang, L. (1997). Epistasis in measured genotypes: *Drosophila* P-element insertions. *Genetics*, 147(1), 157-163.
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., ... & Kaiser, C. A. (2010). The genetic landscape of a cell. *science*, 327(5964), 425-431.
- Guernsey, D. L., Jiang, H., Campagna, D. R., Evans, S. C., Ferguson, M., Kellogg, M. D., ... & Samuels, M. E. (2009). Mutations in mitochondrial carrier family gene SLC25A38 cause nonsyndromic autosomal recessive congenital sideroblastic anemia. *Nature genetics*, 41(6), 651-653.
- de Visser, J. A. G., Cooper, T. F., & Elena, S. F. (2011). The causes of epistasis. *Proceedings of the Royal Society B: Biological Sciences*, 278(1725), 3617-3624.

- Guo, G., & Weiss, B. (1998). Endonuclease V (nfi) Mutant of *Escherichia coli* K-12. *Journal of bacteriology*, 180(1), 46-51.
- Gales, A. C., Gordon, K. A., Wilke, W. W., Pfaller, M. A., & Jones, R. N. (2000). Occurrence of single-point *gyrA* mutations among ciprofloxacin-susceptible *Escherichia coli* isolates causing urinary tract infections in Latin America. *Diagnostic microbiology and infectious disease*, 36(1), 61-64.
- Gammon, K. (2014). Drug discovery: Leaving no stone unturned. *Nature*, 509(7498), S10-S12.
- Gotfredsen, M., & Gerdes, K. (1998). The *Escherichia coli* relBE genes belong to a new toxin–antitoxin gene family. *Molecular microbiology*, 29(4), 1065-1076.
- Ibañez, E., Campos, E., Baldoma, L., Aguilar, J., & Badia, J. (2000). Regulation of Expression of the *yiaKLMNOPQRS* Operon for Carbohydrate Utilization in *Escherichia coli*: Involvement of the Main Transcriptional Factors. *Journal of bacteriology*, 182(16), 4617-4624.
- Hansen, T. F. (2013). Why epistasis is important for selection and adaptation. *Evolution*, 67(12), 3501-3511.
- Hemm, M. R., Paul, B. J., Schneider, T. D., Storz, G., & Rudd, K. E. (2008). Small membrane proteins found by comparative genomics and ribosome binding site models. *Molecular microbiology*, 70(6), 1487-1501.
- Herring, C. D., Glasner, J. D., & Blattner, F. R. (2003). Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. *Gene*, 311, 153-163.

- Hietpas, R. T., Jensen, J. D., & Bolon, D. N. (2011). Experimental illumination of a fitness landscape. *Proceedings of the National Academy of Sciences*, 108(19), 7896-7901.
- Johnson, J. R., Tchesnokova, V., Johnston, B., Clabots, C., Roberts, P. L., Billig, M., ... & Sokurenko, E. V. (2013). Abrupt emergence of a single dominant multidrug-resistant strain of *Escherichia coli*. *Journal of Infectious Diseases*, 207(6), 919-928.
- Kurihara, S., Oda, S., Kato, K., Kim, H. G., Koyanagi, T., Kumagai, H., & Suzuki, H. (2005). A novel putrescine utilization pathway involves γ -glutamylated intermediates of *Escherichia coli* K-12. *Journal of Biological Chemistry*, 280(6), 4602-4608.
- Kim, T., Lee, J., & Kim, K. S. (2013). *Escherichia coli* YmdB regulates biofilm formation independently of its role as an RNase III modulator. *BMC microbiology*, 13(1), 266.
- Kishii, R., & Takei, M. (2009). Relationship between the expression of ompF and quinolone resistance in *Escherichia coli*. *Journal of infection and chemotherapy*, 15(6), 361-366.
- Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., & Mori, H. (2006). Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA research*, 12(5), 291-299.
- Kroll, E. S., Hyland, K. M., Hieter, P., & Li, J. J. (1996). Establishing genetic interactions by a synthetic dosage lethality phenotype. *Genetics*, 143(1), 95-102.
- Lenski, R. E. (1988). Experimental studies of pleiotropy and epistasis in *Escherichia coli*. II. Compensation for maladaptive effects associated with resistance to virus T4. *Evolution*, 433-440.

- Neuhard, J., & Tarpø, L. (1993). Location of the *udk* gene on the physical map of *Escherichia coli*. *Journal of bacteriology*, 175(17), 5742.
- Mulligan, C., Fischer, M., & Thomas, G. H. (2011). Tripartite ATP - independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS microbiology reviews*, 35(1), 68-86.
- Mackenzie, S. M., Brooker, M. R., Gill, T. R., Cox, G. B., Howells, A. J., & Ewart, G. D. (1999). Mutations in the white gene of *Drosophila melanogaster* affecting ABC transporters that determine eye colouration. *Biochimica et Biophysica Acta (BBA)- Biomembranes*, 1419(2), 173-185.
- Marcusson, L. L., Frimodt-Møller, N., & Hughes, D. (2009). Interplay in the selection of fluoroquinolone resistance and bacterial fitness. *PLoS pathogens*, 5(8), e1000541.
- Parera, M., & Martinez, M. A. (2014). Strong Epistatic Interactions within a Single Protein. *Molecular biology and evolution*, msu113.
- Measday, V., Baetz, K., Guzzo, J., Yuen, K., Kwok, T., Sheikh, B., ... & Andrews, B. (2005). Systematic yeast synthetic lethal and synthetic dosage lethal screens identify genes required for chromosome segregation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), 13956-13961.
- Memarian, N., Jessulat, M., Alirezaie, J., Mir-Rashed, N., Xu, J., Zareie, M., ... & Golshani, A. (2007). Colony size measurement of the yeast gene deletion strains for functional genomics. *BMC bioinformatics*, 8(1), 117.

- Costanzo, M., Baryshnikova, A., Myers, C. L., Andrews, B., & Boone, C. (2011). Charting the genetic interaction map of a cell. *Current opinion in biotechnology*, 22(1), 66-74.
- Morozova, O., & Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92(5), 255-264.
- Orr, H. A. (2005). The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, 6(2), 119-127.
- Palmer, A. C., & Kishony, R. (2013). Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nature Reviews Genetics*, 14(4), 243-248.
- Park, S., & Lehner, B. (2013). Epigenetic epistatic interactions constrain the evolution of gene expression. *Molecular systems biology*, 9(1).
- Phillips, P. C. (2008). Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, 9(11), 855-867.
- Phillips, P. C. (1998). The language of gene interaction. *Genetics*, 149(3), 1167-1171.
- Plantinga, T. H., van der Does, C., Badia, J., Aguilar, J., Konings, W. N., & Driessen, A. J. (2004). Functional characterization of the Escherichia coli K-12 yiaMNO transport protein genes. *Molecular membrane biology*, 21(1), 51-57.
- Plantinga, T. H., van der Does, C., Tomkiewicz, D., van Keulen, G., Konings, W. N., & Driessen, A. J. (2005). Deletion of the yiaMNO transporter genes affects the growth characteristics of Escherichia coli K-12. *Microbiology*, 151(5), 1683-1689.

- Silver, L. L. (2011). Challenges of antibacterial discovery. *Clinical microbiology reviews*, 24(1), 71-109.
- Typas, A., Nichols, R.J., Siegele, D.A., Shales, M., Collins, S.R., et al. (2008) High-throughput, quantitative analyses of genetic interactions in E. coli. *Nat Methods* 5: 781–787
- Tenaillon, O., Rodríguez-Verdugo, A., Gaut, R. L., McDonald, P., Bennett, A. F., Long, A. D., & Gaut, B. S. (2012). The molecular diversity of adaptive convergence. *Science*, 335(6067), 457-461.
- Manges, A. R., Johnson, J. R., Foxman, B., O'Bryan, T. T., Fullerton, K. E., & Riley, L. W. (2001). Widespread distribution of urinary tract infections caused by a multidrug-resistant *Escherichia coli* clonal group. *New England Journal of Medicine*, 345(14), 1007-1013.
- Wade, M. J. (2002). A gene's eye view of epistasis, selection and speciation. *Journal of Evolutionary Biology*, 15(3), 337-346.
- Weinreich, D. M. Predicting Molecular Evolutionary Trajectories in Principle and in Practice. *eLS*.
- Weinreich, D. M., Watson, R. A., & Chao, L. (2005). Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*, 59(6), 1165-1174.
- Self, W. T., Hasona, A., & Shanmugam, K. T. (2004). Expression and regulation of a silent operon, *hyf*, coding for hydrogenase 4 isoenzyme in *Escherichia coli*. *Journal of bacteriology*, 186(2), 580-587.

- Whitehurst, A. W., Bodemann, B. O., Cardenas, J., Ferguson, D., Girard, L., Peyton, M., ... & White, M. A. (2007). Synthetic lethal screen identification of chemosensitizer loci in cancer cells. *Nature*, 446(7137), 815-819.
- World Health Organization. (2010). Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 global report on surveillance and response.
- Wright, S. (1932). *The roles of mutation, inbreeding, crossbreeding, and selection in evolution* (Vol. 1, pp. 356-366). na.
- Yu, H. H. Y., Kibler, D., & Tan, M. (2006). In silico prediction and functional validation of σ 28-regulated genes in Chlamydia and Escherichia coli. *Journal of bacteriology*, 188(23), 8206-8212.
- Yao, M., & Kow, Y. W. (1996). Cleavage of insertion/deletion mismatches, flap and pseudo-Y DNA structures by deoxyinosine 3' -endonuclease from Escherichia coli. *Journal of Biological Chemistry*, 271(48), 30672-30676.

Glossary

Compositional epistasis- epistasis resulting from protein-protein interactions

Epistasis- interaction of two genotypes at two or more loci producing a phenotype that is not expected

Functional epistasis- epistatic interactions describing the genetic ordering of regulatory pathways

Statistical epistasis- a deviation of the additive effects of genes, viewed from a population wide scale. Determined by changing to alleles at two loci and comparing the change in fitness to the population mean fitness

Synthetic lethal interaction- result of a combination of mutations that have no or little effect on their own, but together are lethal

Synthetic dosage lethal interaction- similar to synthetic lethal interaction, but when one of the mutations is an overexpression mutation, resulting in a lethal interaction

ORF- Open reading frame, part of the reading frame that contains no stop codons and may be a potential gene.