Université de Montréal

A multiscale framework for microbial evolution to identify the emergence of antibiotic resistance

par Anh-Tien Ton

Département de Biochimie et Médecine Moléculaire
Faculté de Médecine

Mémoire présenté
en vue de l'obtention du grade de M.Sc
en Bio-informatique

Août 2018

Résumé

La résistance aux antimicrobiens (pharmacorésistance) est une crise sanitaire qui menace nos moyens pour contrôler les infections bactériennes. Nos progrès dans la médecine dépendent de nos habiletés à combattre les infections avec des antibiotiques. Ainsi, il est nécessaire de comprendre le mécanisme entourant l'évolution de la résistance aux antibiotiques. Prédire les trajectoires évolutives de la pharmacorésistance demeure une tâche ardue et urgente. Actuellement, notre capacité à prédire les voies d'évolution bactérienne vers la pharmacorésistance est limitée. Il implique de combler plusieurs contraintes sur différents niveaux d'organisation biologique – des propriétés moléculaires des protéines, l'aptitude des organismes et à la dynamique des populations microbiennes.

Dans ce mémoire, je développe un nouveau modèle multiscalaire pour l'évolution microbienne qui intègre principalement la génétique des populations avec la biophysique. Mon système modèle est la ß-lactamase qui concède une résistance contre une large gamme d'antibiotiques ß-lactamines. Tout d'abord, je détermine le paysage d'aptitude de la ß-lactamase en utilisant le balayage mutationnel profond (DMS), un nouvel outil pour tester expérimentalement l'aptitude d'environ 5000 variantes de la ß-lactamase. Ensuite j'intègre ces données expérimentales dans mon modèle computationnel d'évolution microbienne pour étudier les voies évolutives envers la pharmacorésistance.

Dans le premier chapitre, je développe un modèle évolutionniste déterministe combinant la dynamique des populations et les effets biochimiques des mutations pour capturer les effets de la sélection purificatrice avec l'ampicilline. En raison des informations limitées qu'un modèle déterministe peut fournier, dans le deuxième chapitre, je bâtis sur le modèle initial en développant un modèle stochastique de l'évolution microbienne. Ce modèle mis à jour vise à déterminer les mutations qui pourraient être enrichies lors d'un traitement antibiotique. J'étudie également les régimes pour atténuer l'émergence de la résistance. Dans le troisième chapitre, je construis expérimentalement avec le DMS, le paysage d'aptitude de TEM-1 (Temoneira-1), une enzyme de la ß-lactamase pour déterminer son niveau de résistance et sa dépendance envers céfotaxime.

Mots-clés : Résistance aux antibiotiques, évolution, paysage d'aptitude, modèle évolutif, balayage mutationnel profond, coût d'aptitude

Abstract

Antimicrobial resistance is an emerging health crisis that threatens our ability to control bacterial infections. Advances in medical treatments depend on the ability to fight infections with antibiotics. Thus, there is a need to understand the mechanism surrounding the evolution of antibiotic resistance. Predicting the evolutionary trajectories to drug resistance remains a daunting task and is urgently needed. Currently, our aptitude to predict pathways in bacterial evolution to drug resistance is limited. It entails bridging several constraints on various levels of biological organization—from molecular properties of proteins to organismal fitness, to microbial population dynamics.

In this memoir, I develop a new multi-scale framework for microbial evolution that integrates principally population genetics with biophysics. My model system is beta-lactamase that provides broad-spectrum resistance against beta-lactam antibiotics. First, I determine the fitness landscape of ß-lactamase using deep mutational scanning, a novel tool to experimentally assay the fitness of around 5000 variants of beta-lactamase. Then, I integrate this experimental fitness landscape data into my computational model of microbial evolution to study the evolutionary pathways to drug resistance.

In the first chapter, I develop a deterministic evolutionary model combining population dynamics and the biochemical effects of mutations to capture the effects of purifying selection under selection with ampicillin. Due to the limited information that a deterministic model can provide, in the second chapter, I build upon the initial model to develop a stochastic model of microbial evolution. This updated model aims to determine mutations that might be enriched during antibiotic treatment. I investigate the landscape of fitness cost against resistance level. I also investigate drug regimens to alleviate the rise of resistance. In the third chapter, I experimentally determine with DMS the fitness landscape of TEM-1 (Temoneira-1), a ß-lactamase enzyme, to study its resistance level and its dose-dependence for cefotaxime.

Keywords: antibiotic resistance, evolution, fitness landscape, evolutionary model, deep mutational scanning, fitness cost

Table of Contents

List of Tables

List of Figures

List of abbreviations

AMR: Antimicrobial resistance

Amp: Ampicillin

bp: base pair

Cfx : Cefotaxime

DFE: Distribution of Fitness Effects

DMS: Deep mutational scan

DNA: Deoxyribonucleic acid

FC: Fitness Cost

HGT: Horizontal Gene Transfer

LB: Lysogeny broth (Luria broth)

MIC: Mean Inhibitory Concentration

Mut: Mutant

NGS: Next-Generation Sequencing

OD: Optical density (600)

PCR: Polymerase Chain Reaction

RL: Resistance Level

RNA: Ribonucleic Acid

Tau (τ): Time constant

TEM: Temoneira

Tet: Tetracycline

WT: Wild-type

ATP: Adenosine triphosphate

PNK: Polynucleotide Kinase

NFH2O: Nuclease-Free Water

dNTPs: Deoxynucleoside triphosphate

DMSO: Dimethyl sulfoxide

dsDNA/ssDNA: Double-stranded/single-stranded Deoxyribonucleic acid

rpm: revolutions per minute

Acknowledgment

I would first like to thank my supervisor, Adrian Serohijos for giving me the chance to join his group and believing in me and my potential when others did not. I would not be where I am today without his help and guidance. I am grateful to all members of the Serohijos lab for their help throughout my Masters: Yun Zhu and Christopher Savoie for teaching me the experimental aspects of research; and Pouria Dasmeh and Sebastien Boyer for discussing theoretical questions. Without their valuable help, I would have never accomplished what I did during my Masters. I would finally like to thank my family and friends for their support throughout the challenging times endured during my studies. Their support allowed me to push through to complete my work and my memoir.

Chapter 1: Introduction

The discovery of antibiotics is one of the most important medical advances that lead to the reduction of human mortality and morbidity (Andersson et al., 2010). Antibiotic resistance can be described as the ability of microorganisms to survive the drugs designed to eradicate them (Ventola, 2015). Our understanding of the mechanisms behind the evolution of resistance remains incomplete. As such, the emergence of multiple drug resistance in patients has proven difficult to efficiently treat (Higgins, 2007). One possible treatment to alleviate resistance is to reduce the use of antibiotics (Melnyk et al., 2015). This strategy exploits the fitness cost imposed by the evolution of resistance-conferring mutations to purge them out of a population. Thus, the existing approaches to mitigate resistance depend on our understanding of the evolution of resistance.

To gain a better understanding of the evolution of antimicrobial resistance, we developed a stochastic evolutionary model to study the dynamics of resistance-conferring mutations. The stochastic evolutionary model is built by combining theories from biophysics, such as epistasis (Serohijos and Shakhnovich., 2014), and from population genetics, such as effective population size (Charlesworth, 2009). This theoretical model provides valuable insight into the evolutionary pathways to drug resistance. The emergence of resistance is often described as a pathway. This pathway to resistance consists of a set of mutations fixing in a population, leading to resistance (Hall, 2002). Evolutionary pathways are commonly studied by constructing the appropriate fitness landscapes of resistance. Consequently, an accurate construction of the fitness landscape of resistance is critical to increase the accuracy of our models and to study the emergence of resistance.

One of the new methods to construct the fitness landscapes of proteins that confer resistance is deep mutational scanning (Fowler and Fields, 2014). This technique combines library mutagenesis, selection, and next-generation sequencing to assay several thousands of mutational variants. The fitness landscapes constructed from the deep mutational scan can be used with our theoretical models to study the emergence of resistance. In this chapter, we

introduce the concepts of the emergence of antibiotic resistance and how to study the pathways to resistance on fitness landscapes. We also introduce experimental approaches to determine fitness landscapes of resistance.

1. Antibiotic resistance

1.1 Emergence of antibiotic resistance

Antimicrobial resistance has become an emerging problem in healthcare and is among the major causes of mortality and morbidity in clinical settings throughout the world (Guo, 2012). Because of antibiotic-resistant strains, there are fewer antibiotics that can effectively treat infections (Maharjan, 2017). The emergence of antibiotic-resistant strains is partially due to the widespread use of antibiotics in healthcare and agriculture industries (Chang et al., 2015). This increased exposure to antibiotics constitutes a strong and persistent selective pressure for the evolution of resistant strains in a population (Andersson, 2010).

There are also multiple other mechanisms that could lead to the emergence of resistance in a population. The emergence of antibiotic resistance can occur from de novo mutations or other types of genetic changes. The most common mutations for resistance alter antibiotic targets or increase drug efflux rates, but mutations have also been observed to affect gene amplification and to reduce the expression of the antibiotic target (Melnyk et al., 2015). Another important mechanism for a bacterium to acquire resistance is through horizontal gene transfer (HGT) of resistance cassettes between microbes (Sundqvist, 2014, Melnyk et al, 2015). HGT mechanisms include drug modification enzymes, antibiotic target protection, replacement of drug targets, and acquisition of new efflux pumps to clear out the drugs (Andersson and Hughes, 2010, Robicsek et al., 2006). The most common mechanisms for acquiring resistance are presented in **Figure 1**.

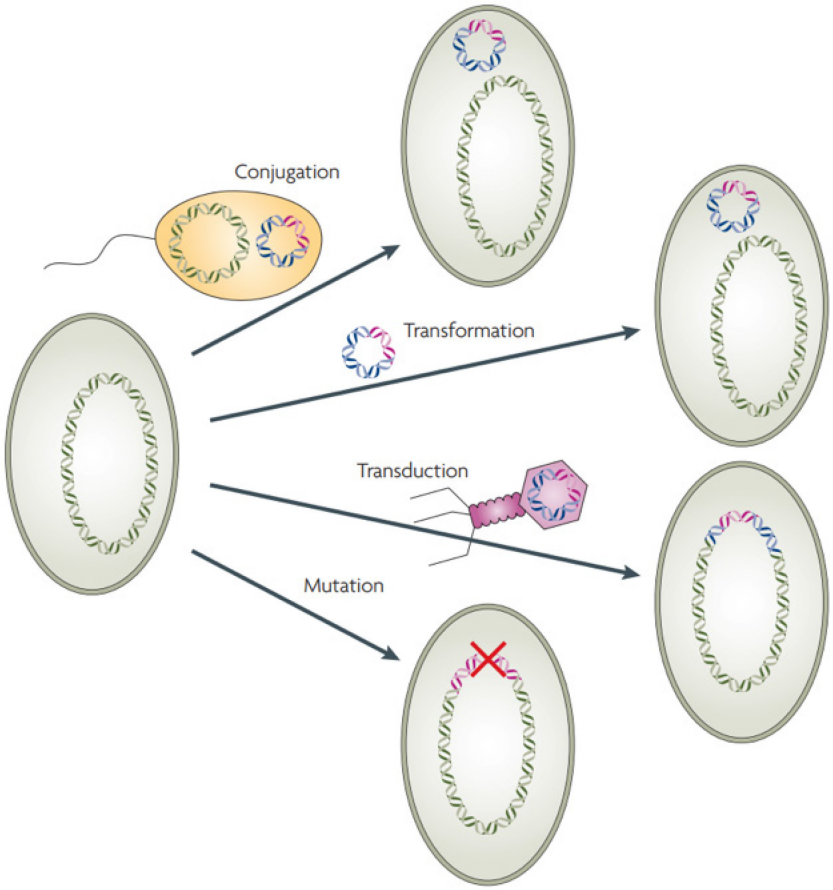

Figure 1. Mechanisms of resistance acquisition.
An antibiotic resistance gene (pink) can be transferred by horizontal gene transfer to another strain by multiple paths: cell-to-cell conjugation, the transformation of DNA, phage-mediated transduction. Resistance mutations can also arise de novo in the new organism. (Adapted from Andersson and Hughes, 2010)

1.2 TEM-1 beta-lactamase

Beta-lactam antibiotics, such as penicillin – the first antibiotic developed, were commonly used to treat bacterial infections. This treatment proved to be efficient until bacteria developed resistance to first-generation beta-lactams antibiotics (Kong et al., 2011). Bacteria developed resistance against beta-lactam antibiotics by producing beta-lactamases. TEM-1, the first beta-lactamase enzyme identified, degrade these antibiotics by hydrolyzing the beta-lactam ring found in numerous beta-lactams (Cooksey, 1990). Thus, TEM-1 can provide a broad-spectrum

resistance against beta-lactams. To counteract the growing problem of TEM-1 producing bacteria, newer beta-lactams antibiotics were developed (Shaik et al., 2015). However, antibiotic resistance emerged against these alternative antibiotics with the appearance of modified beta-lactamases (Shaik et al., 2015). The development of novel antibiotics gave rise to TEM variants with different amino acid sequences that confer different resistance phenotypes (Salverda, 2010). TEM-3, one of the first derivatives of TEM-1, had a single amino acid substitution and is carried by a plasmid which encodes a gene responsible for a new ceftazidime resistance (Shaik et al., 2015). These derivates of TEM-1 likely evolved long ago, but they emerged when faced against these new drugs. Due to a large number of beta-lactam antibiotics, beta-lactamases cover a wide spectrum of resistance and confers good resistance against second, third, and fourth-generation of beta-lactamase inhibitors. Because of the current extensive knowledge and detailed description of the natural evolution of TEM alleles, TEM-1 has been a frequent target to study the emergence of antibiotic resistance.

1.3 Fitness cost of resistance-conferring mutations

Although there have been numerous interpretations of fitness, fitness is generally understood as the ability of organisms to survive and reproduce in their environment. Surviving organisms reproduce and contribute to the gene pool of subsequent generations. Consequently, fitness is a measure of an organism's reproductive abilities and reflects its adaptive capabilities in a specific environment. Following that definition, the fitness of a resistant mutation determines its survival in a population under selective pressure (Orr, 2009). Resistance level is the fitness of a protein under prevailing antibiotic selection environments (Melnyk et al., 2015). Acquisition of a resistance mutation might increase the resistance level, but it can also impose a fitness cost on the organism (Sundqvist, 2014). Fitness cost is defined as the reduced competitive ability of a mutant in the absence of antibiotics (Vogwill et al., 2015). Many antibiotics target important biological processes, and resistance to them imposes large energetic burdens and metabolic costs that reduce competitive ability against antibiotic-sensitive strains (Andersen, 2006; Melnyk, 2015). Because of fitness cost, strains that contain resistant mutations may be depleted in the population when the antibiotic is not present.

It is believed that antibiotic-susceptible strains will competitively eliminate the resistant strains in selection-free environments unless the resistant ones acquire mutations that can compensate for this fitness cost (Maharjan et al, 2017). This fitness loss by the resistant mutants is reflected by numerous changes in functional roles, such as reduced growth rate, reduced transmission rate, higher clearance rate, or decreased invasiveness in the absence of antibiotics (Schulz, 2010). Therefore, resistant mutations are more likely to persist in the absence of antibiotic pressure if they suffer little or no fitness cost. Mutant strains that do not pay a cost or very little cost for their resistance have a higher chance to replace other resistant strains (Johnsen, 2009). Thus, determining the extent of fitness cost and how to modulate it are key to determine the strategies to alleviate the emergence of resistance (Schuzl, 2010).

1.4 Current approaches to alleviate antibiotic resistance

There are several potential strategies to address drug resistance, such as multidrug therapy. Multidrug therapy consists of prescribing a combination of different drugs to a patient when resistance has emerged (Perron et al., 2012). Another approach is by exploiting fitness cost (Sundqvist, 2014). The magnitude of fitness cost is one of the primary factors that increases the development of resistance, but it can also decrease resistance when antibiotic is absent (Guo, 2012). Fitness cost can be exploited to introduce competition between antibiotic-susceptible strains and resistant strains (Lipsitch et al., 2000). Antibiotic-susceptible strains are free from fitness costs and therefore can potentially outcompete resistant strains in a selection-free environment. Exploiting fitness cost by reducing antibiotic use is a viable approach to alleviate resistance.

Nevertheless, reducing antibiotic use once resistance has emerged, has not always been effective at alleviating resistance (Sundqvist, 2010). There are two main hypotheses for the persistence of resistant strains in the absence of antibiotics – genetic linkage, and pleiotropic cost (Andersson, 2010).

1. Genetic linkage is the co-occurrence of resistance genes and other compensatory genes that lead to genetic co-selection. This prevents the elimination of resistance.

Compensatory mutations counteract the effects of fitness costs incurred by resistance mutations. The presence of compensatory mutations might prevent genetic reversions of the resistant genes. Both the resistance genotype and the compensatory genotype individually may confer an overall lower fitness than the combined genotype. Therefore, the resistant-compensated genotype cannot revert back the wild-type genotype (Schulz, 2014).

2. Pleiotropic cost is defined as the effects of ancestor genes on the fitness of resistant mutants (Melnyk et al., 2015). Resistance mutations have highly variable pleiotropic costs. Therefore, they may inherit a fitness that is indistinguishable from their antibiotic sensitive ancestor and become "no-cost" mutations. (Andersson, 2003, Melnyk, 2015). These mutations will persist in an antibiotic-free environment as there is no selective pressure against them.

Fitness cost remains a crucial factor in preventing and reverting resistance development. Newer antibiotics could exploit fitness cost by targeting strains for which resistance mechanisms confer a high fitness cost and a low compensation system (Andersson, 2010). There is a need to better understand the biological mechanisms behind fitness cost and how to incorporate them in our quantitative models of antibiotic resistance evolution.

2. Pathways to resistance on fitness landscapes

2.1 Hierarchy of constraints on the pathways to drug resistance

Mutational pathways to adaptation, including antibiotic resistance, are influenced by numerous constraints at distinct levels of biological organization, as presented in **Figure 2**. However, there is an incomplete understanding of the pathways to antibiotic resistance as these constraints are traditionally studied separately (Serohijos et al., 2014; Harms and Thornton., 2013; Wilke et al., 2012). The mutational pathways to resistance are defined by the mapping between microbial fitness and the molecular properties of the target gene. This mapping is referred to as the fitness landscape.

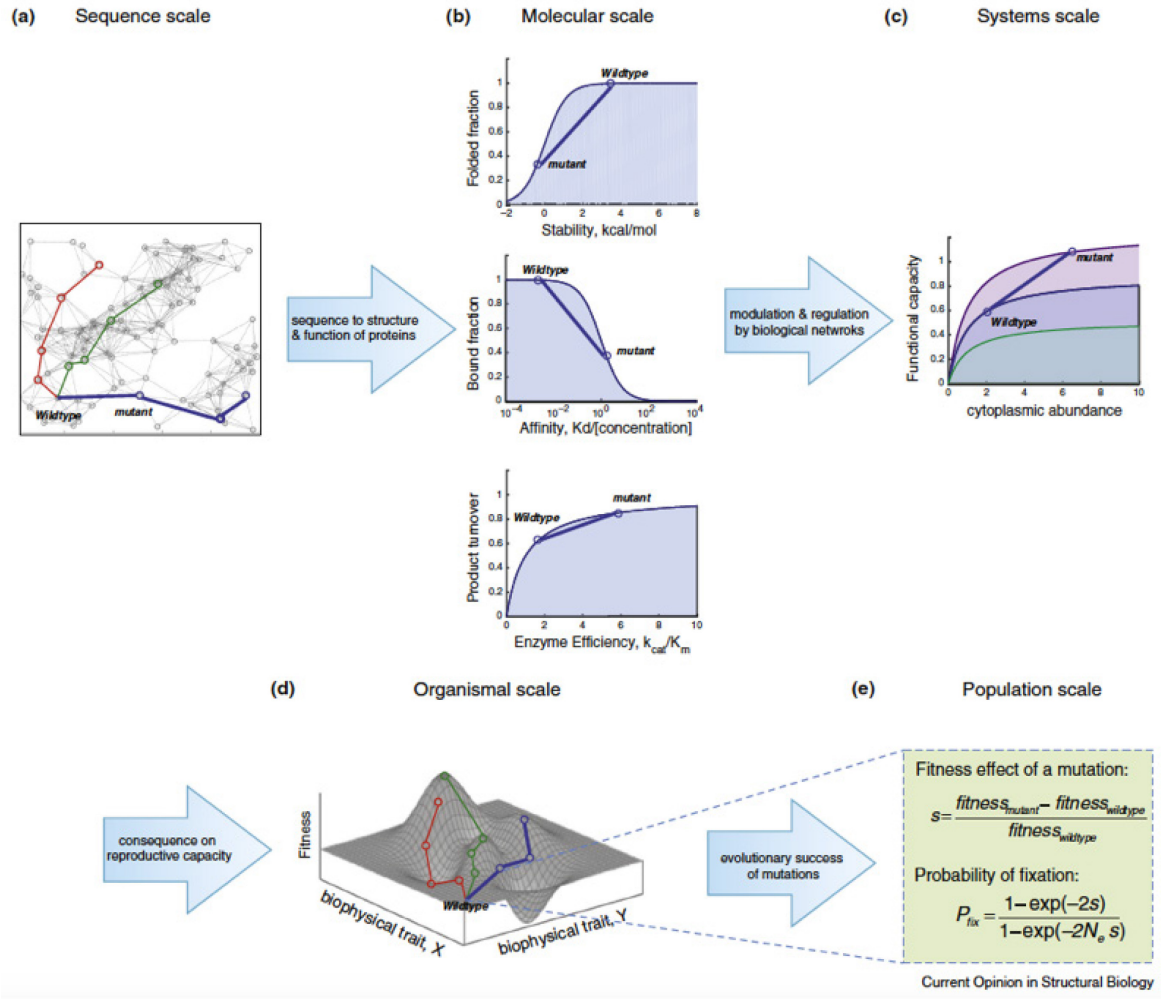

Figure 2. Hierarchy of constraints in evolution.
Combining fitness with the biophysical properties of macromolecules will better shape the genotype-phenotype map (fitness landscape). **A.** Effect of sequence variation on the phenotype. **B.** The relationship between biophysical traits and fitness effects of the organism. **C.** The effects of mutations are regulated by biological networks. **D.** Fitness landscape built on biophysical properties. **E.** Importance of population genetics in the construction of fitness landscapes, such as the selection coefficient and the probability of fixation. (Adapted from Bershtein et al., 2016).

2.2 Fitness Landscapes and the Genotype-Phenotype relationship

The genotype-phenotype map, referenced as the fitness landscape, visualizes the relationship between genotype and reproductive success (Visser et al., 2014). This landscape may be represented as a function, where the x-y planes are genotypes and the z-axis is fitness. The

fitness could be any measure of reproductive success (e.g. growth rates) or any phenotypes (e.g. drug resistance). Such maps define feasible combinations of mutations ("the mutational pathways") towards drug resistance (Hartl, 2014). These evolutionary pathways are dictated by the molecular mechanisms revealed by the landscape. A sample fitness landscape with the evolutionary pathways for the evolution of cancer is presented in **Figure 3**.

Consequently, evolutionary pathways are represented as *walks* on the fitness landscape. Walks around the map represent evolution with small changes to fitness, and *climbs* on top of fitness peaks represent adaptation and increased fitness (Visser et al., 2014; Wright, 1932). A landscape based on resistance can be used to predict paths to fitter genotype or resistance (Weinreich, 2006). However, apart from knowing the fitness landscape, several other constraints that affect the mutational pathways to drug resistance need to be taken into consideration. These include epistasis and the population dynamics of the evolving microbial population. Previous theoretical evolutionary models have considered the importance of accounting for these constraints, and have been successfully implemented to study evolution. Therefore, the inclusion of these parameters is not a novelty in the field of evolutionary models. However, these approaches have not been used to study emergence of antibiotic resistanc. Current evolutionary models will be further discussed in section 2.5.

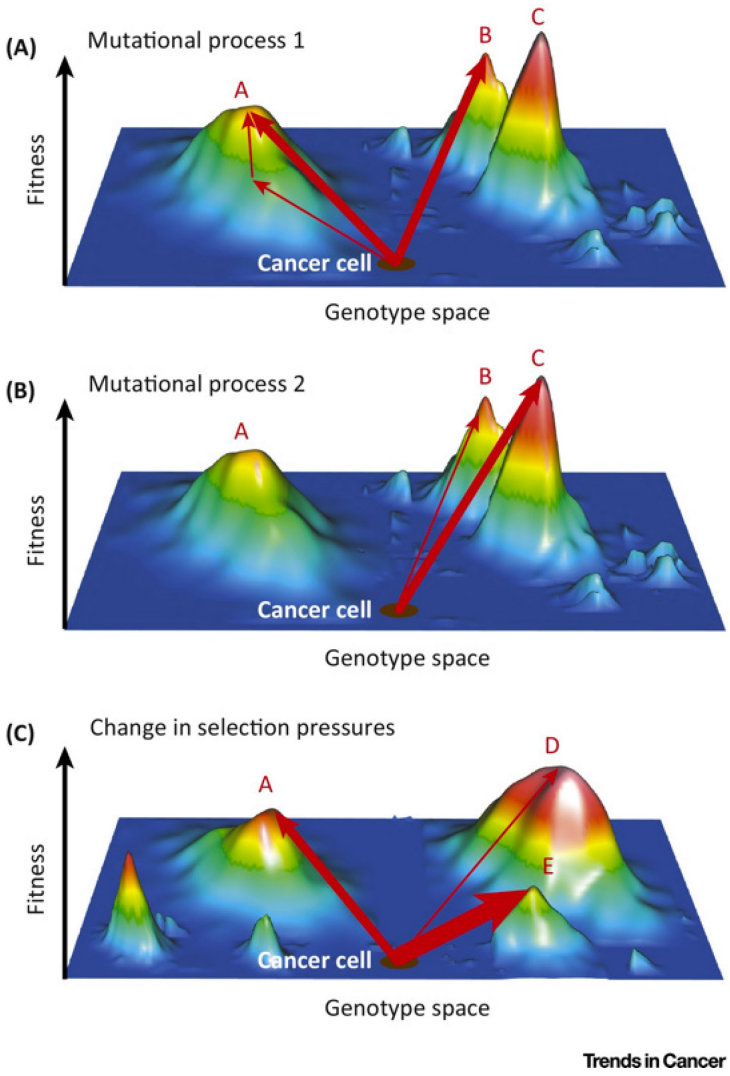

Figure 3. Evolutionary pathways on the fitness landscapes.
Evolutionary pathways of resistant mutations are influenced by the shape and accessibility of fitness peaks on the landscape. The vertical axis represents the fitness level of all genotypes. **A.** A specific set of mutations is allowed to climb to peak A or B which confers resistance. **B.** A different set of mutations which are constrained by biophysical traits or population genetics are only allowed to move to peak B or C. **C.** A change in selective pressure such as a different antibiotic, shifts the fitness landscape and new resistance peaks are formed. Although all pathways in A, B, and C reach resistance, they are constrained to take a different set of mutations (pathways) to reach a fitness peak. (Adapted from Lipinski et al., 2016).

2.3 Epistatic constraints on the pathways to drug resistance

Protein biophysics properties such as folding stability, solubility, and maintenance of functions, impose constraints on the evolutionary pathways leading to drug resistance (Serohijos and Shakhnovich, 2014). However, predicting and reproducing the pathways to resistance is much more complex due to molecular epistatic constraints. Epistasis is broadly defined as the composition of a genotype and the influence that this specific genetic background has on a set of alleles (Phillips, 2008). Consequently, the phenotypic effect of a mutation at one genetic site can change, depending on which alleles are present at other sites. The mutations, in this case, depend on the evolutionary background of the population, and epistasis can greatly influence evolution in populations. Therefore, for some of the resistance-conferring mutations, they are dependent on the genetic background from which they arise. Importantly, epistasis suggests that the order of mutations in the evolutionary pathway is crucial, and the effects of mutations are not always simply additive (Weinreich, 2006). The ruggedness and curvature of the fitness landscape influence epistatic effects on the pathways; therefore, the population may need to traverse low fitness valleys to reach fitness peaks. Consequently, accurately identifying the role of molecular epistasis in evolution is critical to determine the potential pathways taken by populations on the fitness landscape (Breen et al., 2013).

2.4 Population dynamics constraints on the pathways to resistance

Population size is another major constraint on the types of mutational pathways that can arise in the evolution of antimicrobial resistance. Population size (N_e) is a core concept in both evolutionary biology and population genetics as it tunes the balance between selection and mutational drift (Charlesworth, 2009). The force of selection is proportional to the population size. In large populations, mutations with slightly beneficial or deleterious effects can reach fixation or escape genetic drift, thus allowing the mutants to outcompete the wild-type. In small populations, only mutations with strong selection coefficients will fix in the population, as the effects of genetic drift are significantly more intense. In the case where there is only a single individual in a population (N_e=1), there is no competition and there is no selection. All

mutations, deleterious or beneficial, can be fixed. The role of population size in the probability of fixation is formalized by the classic Kimura formula, shown in **Figure 2E**, derived for the simple case of a monoclonal population (Kimura, 1968).

In general, the following forces need to be considered for the evolution of a biological system:

i. *Natural selection* is the process in which alleles for fitter organisms become more frequent in a population as they survive and reproduce, consequently transmitting their alleles to the next generation (Andrews, 2010). The selection coefficient *s* is often used to quantify the differences in fitness between different genotypes (Hartl et al., 2007).
ii. *Mutation* is the process in which new alleles appear in the genome. Mutations can either be deleterious, beneficial, or neutral. Deleterious mutations result in the death or deformation of the organism, while beneficial mutations profit the organism by making it fitter. Mutations are not necessarily a binary process, but instead cover a wide distribution as there exist other mutations, such as neutral mutations (Eyre-Walker et al., 2007).
iii. *Genetic drift* is the process in which allele frequency is changed due to random sampling. Genetic drift has a higher impact on smaller population size, as the involved allele is either lost or fixed in the population at a faster rate (Kliman et al., 2008). The level of variability in a population and the effect of selection relative to drift can be determined through the effective population size. (Charlesworth, 2009).

2.5 Current development of theoretical evolutionary models

Presently, there exist numerous evolutionary models integrating the various molecular, biophysical and population dynamics constraints to provide valuable insight into the process of evolution. Sailer and Harms developed an evolutionary model considering epistatic constraints (Sailer and Harms, 2017). They studied the role of high-order epistasis by creating a model that would remove epistasis from fitness landscapes. They generated two sets of evolutionary pathways, ones from the fitness landscapes with epistasis, and ones from their computational fitness landscapes without epistasis. Using their model, they determined that the accessibility and probability of evolutionary trajectories were affected by the magnitude of

epistasis, not the order itself. Interestingly, these findings were contrary to previous experimental work performed by Weinreich (2006). Nonetheless, they arrived at the same conclusion from their simulations that epistasis and interactions between mutations strongly shape the evolutionary trajectories on fitness landscapes.

Another model has been developed by Meyer and Wilke, integrating protein structure and sequence variation constraints (Meyer and Wilke, 2013). Their model was used to identify resistance sites in neuraminidase under oseltamivir selection. They demonstrated that using structural information with protein sequences can be a powerful predictive tool to identify sites of interest in resistance. They observed that the accuracy of their model for specific proteins improved significantly by considering the biophysical aspect of solvent exposure. Their model provided additional accurate evolutionary information on the resistance mutation sites.

There also exist other evolutionary models integrating multiple constraints simultaneously. The model combining protein biophysics and population dynamics constraints (Serohijos and Shakhnovic, 2014) was used to determine the effects of selection for protein folding stability on the patterns of evolutionary forces in coding regions. In a similar fashion, a model combining population genetics with simple biophysical protein folding (Wylie and Shaknovich, 2011) was used to study the interplay between biophysical and population genetic forces on the shape of the distribution of fitness effects.

Therefore, although there exist a few evolutionary models combining the various molecular and population genetics constraints, we aim to integrate these parameters as a novel approach to specifically elucidate the emergence of resistance on fitness landscapes.

3. Stochastic evolutionary models of population genetics

Our ability to study the genetics of populations depends on our capacity to accurately construct models that capture the essential biological features of populations. A deterministic model of evolution was implemented by Dieckmann and Law in 1996. They identified a deterministic approximation of the stochastic evolutionary process and demonstrated that

evolutionary dynamics can be represented as directed random walks on adaptive landscapes. From their model, they showed that evolutionary paths are equivalent to hill-climbs on landscapes (Dieckmann and Law, 1996). However, their models provide limited insight into the process of mutant selection in a population, as advantageous mutations will always fix in a deterministic model (Nowak, 2006). This is not always the case in nature. In stochastic models, beneficial mutations are not always guaranteed to reach fixation, and there is always a risk of extinction for the mutant. Stochastic models can, therefore, provide additional understanding of actual population evolution.

There are two primary stochastic population genetic models used throughout most studies: The Wright-Fisher model and the Moran model. The Wright-Fisher model represents a population of non-overlapping generations while the Moran model represents a population of overlapping generations (Wakeley, 2008). A population is said to evolve under the Moran process if it satisfies three conditions: the population remains at a constant size N; generations can overlap; and at discrete time intervals, two individuals are chosen randomly to undergo the birth-death process (Nowak, 2006). The model of interest to this memoir is the Moran model. It can replicate the effects of neutral drift, and the probabilistic dynamics of two alleles competing for dominance at populations of finite size (Wakeley, 2008). Studies implementing the stochastic Moran process to elucidate evolutionary dynamics has already been published previously (Muirhead and Wakeley, 2009; Harper and Fryer, 2016). Their Moran model approaches provide the means of modeling a wide variety of fitness schemes. They demonstrated that the Moran model is not strongly dependent on approximations, and therefore this stochastic process can be applied to study evolution requiring population parameters such as selection and effective population size. Thus, we will be using these principles in constructing our stochastic evolutionary model. It will be used to study the emergence of resistance in populations. Our Moran model based on population parameters will be further described in Chapter 4.

4. Experimental approach to determine fitness changes

4.1 Distribution of Fitness Effects

Antibiotic pressure imposes changes on a population and selects for diverse types of mutants in a population, such as resistant mutants. The Distribution of Fitness Effects (DFE) is defined as the spectrum of changes on the fitness of an organism upon genetic perturbations (Eyre-Walker, 2007). The DFE can be used to study the range of mutations observed in a selective environment (Martinez et al., 2000). The DFE is mainly composed of three mutations: beneficial, neutral, and deleterious. Beneficial mutations in the DFE are of special interest as they can potentially confer resistance.

Beneficial mutations are the rarest ones, but they also have the highest impact on the fitness of the organism (Silander et al., 2007). When these mutations are selected for fixation, they contribute substantially to adaptive evolution and reduce genetic diversity (Eyre-Walker, 2006). The small number of beneficial mutations shapes the DFE into an exponential tail (Orr, 2003). The DFE becomes increasingly like an exponential distribution as the population evolves, and strongly advantageous mutations are selected for fixation (Sanjuan et al., 2004, Eyre-Walker et al., 2007). The exponential tail of the DFE for beneficial mutation is represented in **Figure 4**. The spectrum of selective effects defined by the DFE is critical to determine the type and frequency of mutations fixing in the population. Due to its importance, there are multiple experimental methods to quantify the DFE, such as deep mutational scan – defined in the next section.

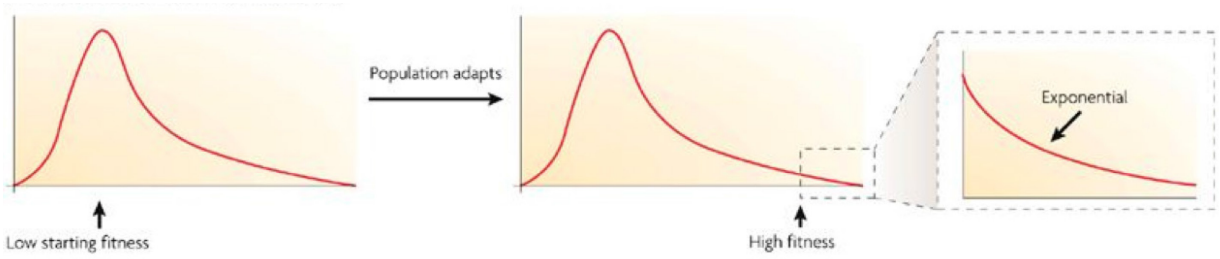

Figure 4. Distribution of fitness effects of new mutations.
The population (represented by the black arrow) starts at a state of low fitness and evolves to a state of high fitness. As the DFE remains constant through evolution, the DFE for advantageous mutations available to the population becomes exponential towards the right-hand tail of the distribution. (Adapted from Eyre-Walker, 2007).

4.2 Deep mutational scan

Deep mutational scanning (DMS) is a new high-throughput approach to comprehensively measure a fitness landscape. It combines saturation mutagenesis, selection, and high-throughput DNA sequencing to assay the functional effects of several thousands of mutations on a protein (Fowler et al., 2010; Fowler et al., 2014). DMS can be used to produce a high-resolution, fine-scale map of protein sequence-function relationships (Fowler et al., 2014). The framework for a DMS is presented in **Figure 5.** This approach generates large-scale mutagenesis datasets containing a functional score for each mutation variant. A large dataset of mutations renders the task of profiling their functional effects on proteins much more manageable and eases the prediction of mutation progression. By performing DMS in the presence of a drug, it is possible to generate a nearly complete map of resistance to the specific drug (Fowler et al., 2010). The map is useful to guide the development of drug treatments to alleviate the emergence of resistance.

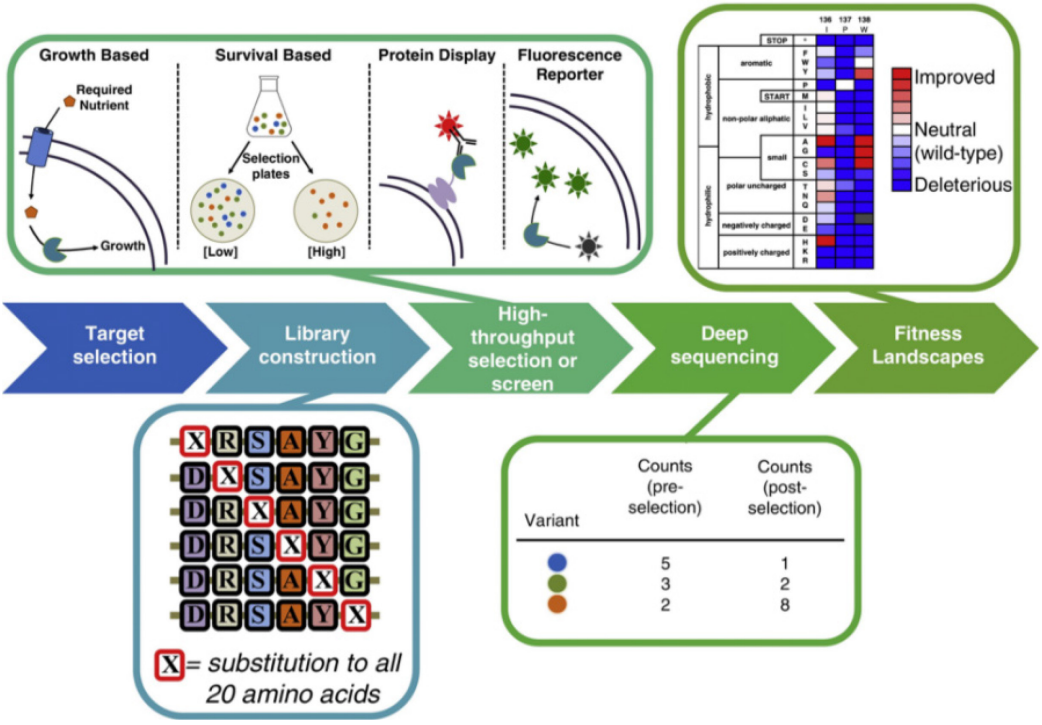

Figure 5. Overview of deep mutational scanning.
Deep mutational scanning can be used to assess the functional consequences of all mutant variations in a protein. A comprehensive mutagenesis library of a protein of interest is constructed. The library is subjected to high-throughput selection or screen for function. Variants are counted via deep sequencing. Fitness landscapes are built from the normalized functional score obtained from sequencing. (Adapted from Wrenbeck et al., 2017).

Chapter 2: Problem and Specific Aims

Problem

Understanding the emergence of resistance in microorganisms is an urgent problem in medicine and public health. Due to resistance, the inefficacy of first-line and second-line antibiotic treatments is forcing healthcare specialists to use stronger drugs that may potentially be toxic to patients (Ventola, 2015). The rise of resistance prevents the effective treatment of common infectious diseases caused by bacteria, parasites, viruses, and fungi, resulting in prolonged illness, or even death. Therefore, there is a need to predict and identify the emergence of resistance. Currently, determining the first passage time of resistance by analyzing the evolutionary trajectories on fitness landscapes might be possible, as demonstrated by numerous groups (Hartl, 2014; Palmer et al., 2013; Poelwijk et al., 2007; Martinez et al., 2007; Rodrigues et al., 2016). Here, we define the first passage time of resistance as the event where the stochastic process of evolution encounters its first resistance-conferring mutation. The emergence of resistance is often described as a specific series of mutations an organism acquires. If we can retrace those trajectories, we could potentially identify the first mutations arising in resistance.

There are only a few studies that combine biophysics and population genetics to study evolution (Wylie and Shakhnovich, 2011; Serohijos and Shakhnovic, 2014). These approaches have certain limitations due to the robustness of genotype-phenotype relationships and quantitative disagreements between models and experiments. Although these limitations hamper model accuracy and predictivity, solving these limitations is outside the scope of this memoir. Nonetheless, these models provide valuable approximations to study evolution, but they have not been applied to study the emergence of resistance. Consequently, the application of evolutionary models and theories to the resistance problem is largely unknown. The role and contribution of molecular biophysics and population genetics to the emergence of resistance remains unclear. Therefore, elucidating their role would be greatly beneficial to our capacity to fight resistance. Integrating both distinct and non-interacting disciplines could provide a better understanding of the evolutionary trajectories to resistance. In this memoir,

we are using concepts from theoretical evolutionary studies to investigate the real-world crisis of drug resistance.

Aims

In this memoir, we will combine techniques from biophysics and biochemistry with principles from evolutionary biology and population genetics to develop a framework to study drug resistance. This framework will combine theory and experiment to predict the near-term evolution of resistance against β-lactamase enzymes, the primary targets of several antibiotics. One of the core concepts borrowed from previous evolutionary models is the principle that fitness landscapes (genotype-phenotype mapping) and evolutionary pathways can be used to study the progress of a population towards resistance. Therefore, we aim to use fitness landscapes with the evolutionary models derived from population genetics to construct the framework of a new multiscale evolutionary model used to predict the emergence of resistance. The models can also be used to further establish the link between experimental fitness landscapes and theoretical evolutionary models.

1. **The first objective is to develop an evolutionary model that integrates principally population genetics constraints.** We first determine the possibility to relay the initial passage time to resistance acquisition on the fitness landscape. More importantly, we seek to integrate the theoretical work on evolution to experimental results of fitness landscapes.
2. **The second objective consists of determining the contribution of fitness cost on the survivability – the ability to remain alive, of resistant mutants**. Using the evolutionary model developed in the first objective, we study the role and the clinical relevance of fitness cost in resistance. We determine the relationship and dependency between fitness cost and resistance level.
3. **The third objective is to determine the fitness landscape of TEM-1 under cefotaxime selection.** By constructing the fitness landscape of TEM-1, we can study the resistance level and dose-dependence for cefotaxime. Cefotaxime is a β-lactam

antibiotic and is part of the third-generation class of cephalosporins. It is often used to treat penicillin and penicillin-derivate resistance (Ma et al., 2002). Therefore, our fitness landscape can be used to test the evolutionary model and to study the emergence of cefotaxime resistance in TEM-1.

The development of a multi-scale framework for microbial evolution will provide insight into the quantitative relationship between the biochemical properties of the target gene, selection regimes, population dynamics, and, more importantly, the first passage time to drug resistance.

Chapter 3: A deterministic evolutionary model for purifying selection

In this chapter, we determine the potential pathways for the emergence of resistance in TEM-1 β-lactamase. We relate the biochemical effects of mutation on target genes with the population demography and the selection regime (drug dosage) experienced by the bacteria. We propose a deterministic evolutionary model of additive fitness. We use the model to investigate the underlying molecular mechanism for the emergence of resistance. We determine if the selection coefficients of resistant mutants are sufficient to recapitulate the beneficial mutations observed in the literature. Using the model, we found that in the context of purifying selection, comprehensive fitness landscapes capture a large fraction of enriched mutations retrieved from the literature. Nonetheless, our deterministic model of additive fitness does not fully capture the pathways to resistance.

Methodology

Fitness landscape of TEM-1 β-lactamase under Amp selection

Our deterministic evolutionary model requires fitness landscapes as input. To construct the comprehensive fitness landscape of TEM-1 β-lactamase under ampicillin (Amp) selection, we used the DMS dataset from Stiffler et al. (2015). Their dataset consists of the relative fitness effects for all single amino acid mutations in TEM-1 under selection for several concentrations of Amp. To generate this dataset, they constructed a whole-gene saturation mutagenesis library of TEM-1. Their library was screened on a growth and survival basis in different selection conditions, followed by next-gen sequencing. The relative fitness for each TEM-1 variant is defined as:

$$F_i^a = \log_{10}\left(\frac{N_i^{a,sel}}{N_i^{a,unsel}}\right) - \log_{10}\left(\frac{N_i^{wt,sel}}{N_i^{wt,unsel}}\right) \qquad \text{(Eq. 1)}$$

where the relative fitness F_i^a of each amino acid mutation a at each position i is determined as the logarithm in the allele counts (N) between the selected population ($N_i^{a,sel}$) and the unselected population ($N_i^{a,unsel}$), relative to the wild-type allele. Their DMS of TEM-1 under Amp selection is presented in **Figure 6.**

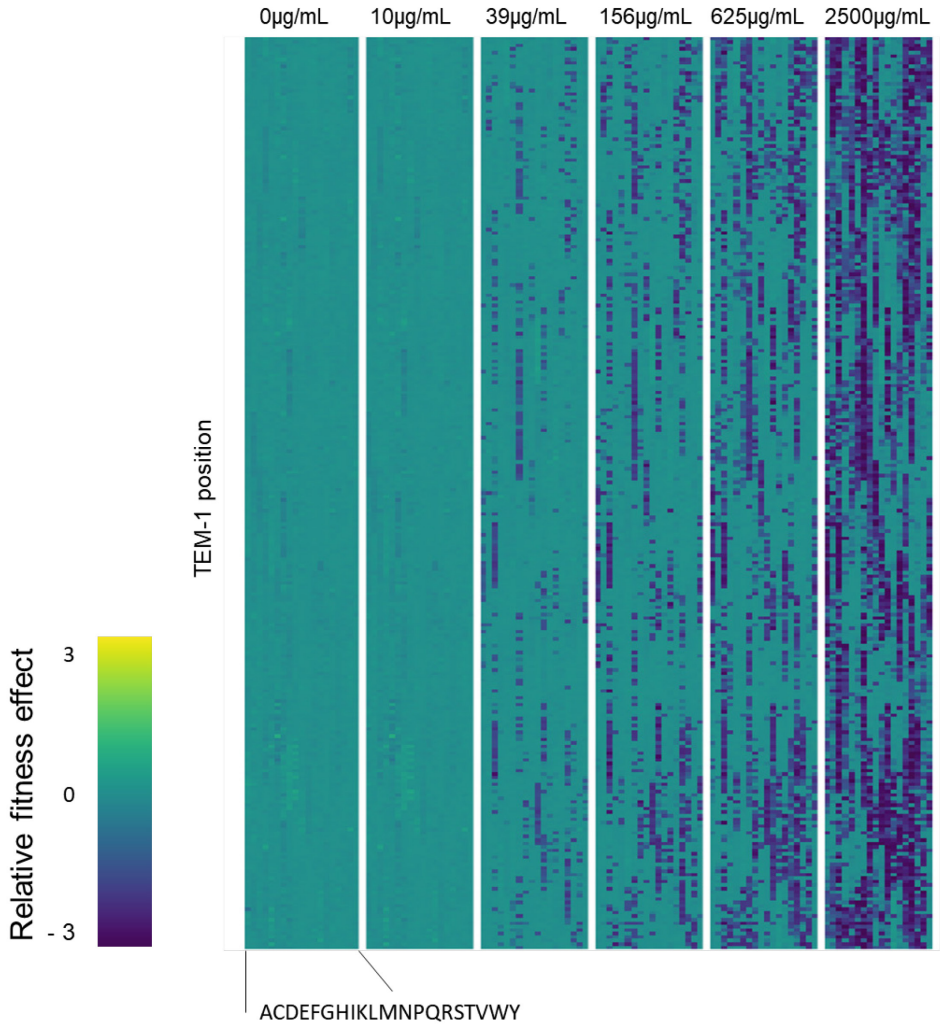

Figure 6. Comprehensive fitness landscape of β-lactamase under Amp selection. Each DMS represent the relative fitness of mutants in specific Amp selection conditions. Each row represents a TEM-1 position, and each column is one of the twenty possible amino acid mutation. Beneficial mutations are highlighted in yellow while deleterious mutations are highlighted in blue. We use the 2500 µg/mL Amp set for our simulations. (Adapted from Stiffler et al., 2015).

Deterministic evolutionary model algorithm for purifying selection

The model consists of a computational workflow to replicate evolutionary pathways on a fitness landscape. We simulated the evolution of the TEM-1 library under constant selection at 2500 μg/mL Amp. The effective population size N_e was set to 10^6 individual cells. The model replicates purifying selection as we iteratively remove deleterious alleles from the pool of available mutations. An initial protein variant with low fitness is selected as the starting point for the simulations and is defined as $F_{current}$. Another mutation is then randomly selected from the dataset as a mutational attempt and is defined as $F_{attempt}$. To determine if the mutation was fixed and retained, we calculated its selection coefficient with Eq. 2:

$$s = F_{attempt} - F_{current} \qquad \text{(Eq. 2)}$$

where the fitness for the initial strain is $F_{current}$ and the fitness for the new mutant is $F_{attempt}$. As the model assumes simple additivity of fitness effects, it does not take epistasis into consideration.

We calculated the probability of fixation (P_{fix}) for the mutational attempt using Kimura's formula:

$$P_{fix} = \frac{1-\exp(-2s)}{1-\exp(-2N_e s)} \qquad \text{(Eq. 3)}$$

Where s is the selection coefficient for the mutational attempt, and N_e is the effective population size. We determine if the mutation was fixed by comparing P_{fix} to a random number generated between 0 and 1. If P_{fix} is larger than the random number, then the mutation is fixed. The fitness of $F_{current}$ is updated to the fitness of the new mutant. If P_{fix} is smaller than the random number, then the mutation is discarded, and the iteration moves to the next mutational attempt. Although N_e inherently accounts for the stochasticity of evolution, we used the P_{fix} step to ensure that only strongly beneficial mutations are fixed in the population, thus maintaining the additive fitness characteristic of our model. A schema of the full algorithm is shown in **Figure 7**. The evolutionary process is repeated until convergence where the percentage increase in relative fitness between fixed mutations is smaller than 5%, or until

a total of 1000 mutational attempts have been performed. As there is only a pool of 4997 mutations available and the model assumes additivity of fitness effects, 1000 mutational attempts were deemed sufficient for the model to capture the dynamics of evolution.

The algorithm keeps track of several variables for each trajectory: the fitness values of each mutational attempt; all fixed mutations occurring in a pathway; and the selection coefficient of each fixed mutation. The evolutionary model was written in Python, and simulations were performed on an iMac, version 10.12.6 with 3.2 GHz Intel Core i5, and 16GB memory.

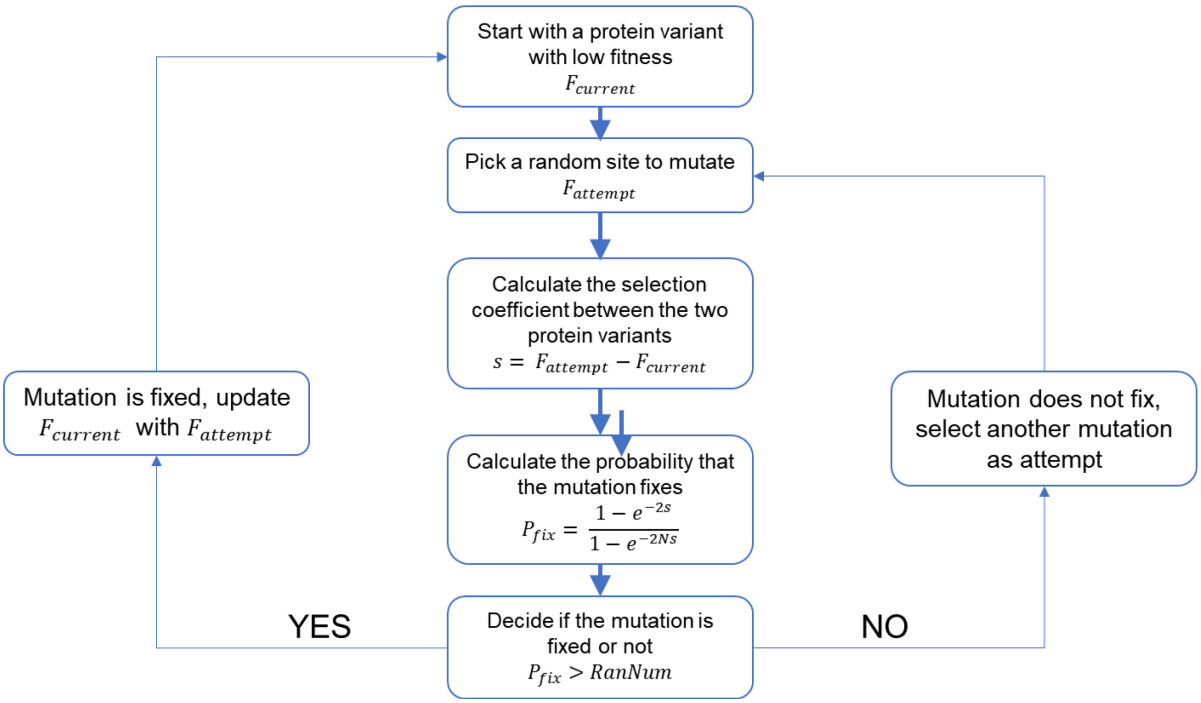

Figure 7. Algorithm for the deterministic evolutionary model.
With the available fitness landscape, we build the model of microbial evolution that replicates bacterial population dynamics to study the emergence of resistance. We calculate the selection coefficient and the probability of fixation as defined by Kimura, to determine which mutations are fixed in our simulations. Beneficial mutations will have a higher probability to fix than a deleterious mutation. We use the algorithm to simulate 10^6 evolutionary trajectories to study the emergence of resistance.

Analysis of the evolutionary trajectories

The population dynamics simulations generate 10^6 pathways, each with their set of fixed mutations. In our model, each step of an evolutionary trajectory/pathway is defined by the selection coefficient of its fixed mutations. Therefore, we calculated the overall mean and median trajectory of all the pathways in the simulations to determine if they recapitulate the evolutionary behavior. In case the mean and median did not recapitulate the evolutionary behavior, we clustered the trajectories to determine their different evolutionary behaviors. We determined an optimal sample size to perform cluster analysis. The sample size can be used to estimate the overall behavior of the pathways. The sample size (n) is defined as:

$$n = \frac{z^2 p(1-p)}{M^2} \qquad \text{(Eq. 4)}$$

Where z is the z-score for the confidence level, p is the sample proportion, and M is the margin error. To ensure that we capture the behavior of the pathways, we set the confidence interval at 95% with a z-score at 1.96; p at 0.5 to ensure the largest sample size calculation, and M at 0.01 for a margin of error of 1%. Using these parameters, we obtained a sample size of 9604 pathways. For the cluster analysis, we randomly sampled 10,000 pathways from the complete set. The sampled trajectories are grouped using the *kml* package (Genollini C. et al., 2015). The *kml* algorithm was set to generate a number of clusters to optimally represent the evolutionary pathways. We calculated the centroid and the standard deviation for each cluster. We performed the sampling and clustering five times, and the average centroid and standard deviation are used to assign all 10^6 evolutionary trajectories to their respective cluster. The analysis was performed with R 3.3.0 (R Core Team, 2013).

Comparison of fixed mutations from simulation with mutations enriched from clinical isolates and experimental laboratory evolution

The mutations from clinical isolates and laboratory evolution under purifying selection are described in **Table 1**. The sets of enriched mutations under Amp selection in laboratory evolution are obtained from Bershtein et al. (2008), and the clinical isolates mutations under Amp selection were obtained from Hall (2002), Matagne et al. (1998), Salverda et al. (2010);

Brown et al. (2010); and Imtiaz et al. (1994). The mutations from the literature were compared against mutations observed in our simulations.

Clinical isolates	Clinical isolates (cont.)	Enriched mutations
A42G	A184V	N52D
I47V	T188I	N52S
M69L	I208M	E63A
M69V	E240K	K111E
G92S	R241H	K111R
M96I	R244S	R120G
E104K	R244C	E147G
S130G	R244T	H153R
R164C	T265M	M182T
R164H	S268G	L201P
R164S	R275L	I208L
W165R	R275Q	I208V
M182T	N276D	K288R
A184V		

Table I. Clinical isolates and enriched mutations for TEM-1 under Amp selection.

The mutations were sampled from the literature on TEM-1 resistance. Clinical isolates represent mutations that were observed in a clinical setting, while the enriched mutations correspond to mutations that were observed in laboratory evolution. The mutations are used to compare against our results from the simulations. The mutations are presented in alphabetical order.

Results

Evolution trajectories form clusters

We build a deterministic evolutionary model of additive fitness. The model generates distinct evolutionary trajectories. We observe that the majority of the trajectories reach resistance fitness within ten mutational attempts or less. By analyzing the dispersion of the trajectories,

we observe that the overall mean and median of the trajectories reach an optimum with diminishing returns in fitness, as shown in **Figure 8A**. Therefore, the average and median of the evolutionary trajectories do not reflect the complex evolutionary dynamics, as they do not account for different growth behaviors.

Instead, the evolutionary trajectories form distinct clusters. We use K-means clustering to group the pathways. We identify three cluster centroids from the evolutionary trajectories. The first cluster represents ~43% of the pathways and corresponds to “fast” adaptation whereby resistance was acquired using only one mutation (**Figure 8B**, red). The second cluster represents ~29% of the pathways and corresponds to “intermediate” adaptation whereby resistance was acquired in two mutations (**Figure 8B**, blue). The third cluster represents ~28% of the pathways and corresponds to “slow” adaptation whereby resistance was acquired in three or more mutations (**Figure 8B**, green).

Next, we investigate the selection coefficient of the mutation or their combination for the three clusters (**Figure 8C)**. We define a mutational "step" as a fixed mutation. Trajectories in the fast adaptation cluster required only one step in average to reach resistance. The first step in the fast adaptation cluster (Cluster #1) is dominated by mutations with high selection coefficients that confer large fitness gains. Subsequent mutational steps confer minimal gains in fitness as the trajectories already reached a fitness plateau. For the intermediate adaptation cluster (Cluster #2), the first and second steps cover a broader and wider range of selection coefficients compared to Cluster #1. Thus, their density is lower compared to the fast adaptation cluster as they are not grouped around the same selection coefficients. The slow adaptation cluster is characterized by a high number of mutational steps that cover a broad range of selection coefficients.

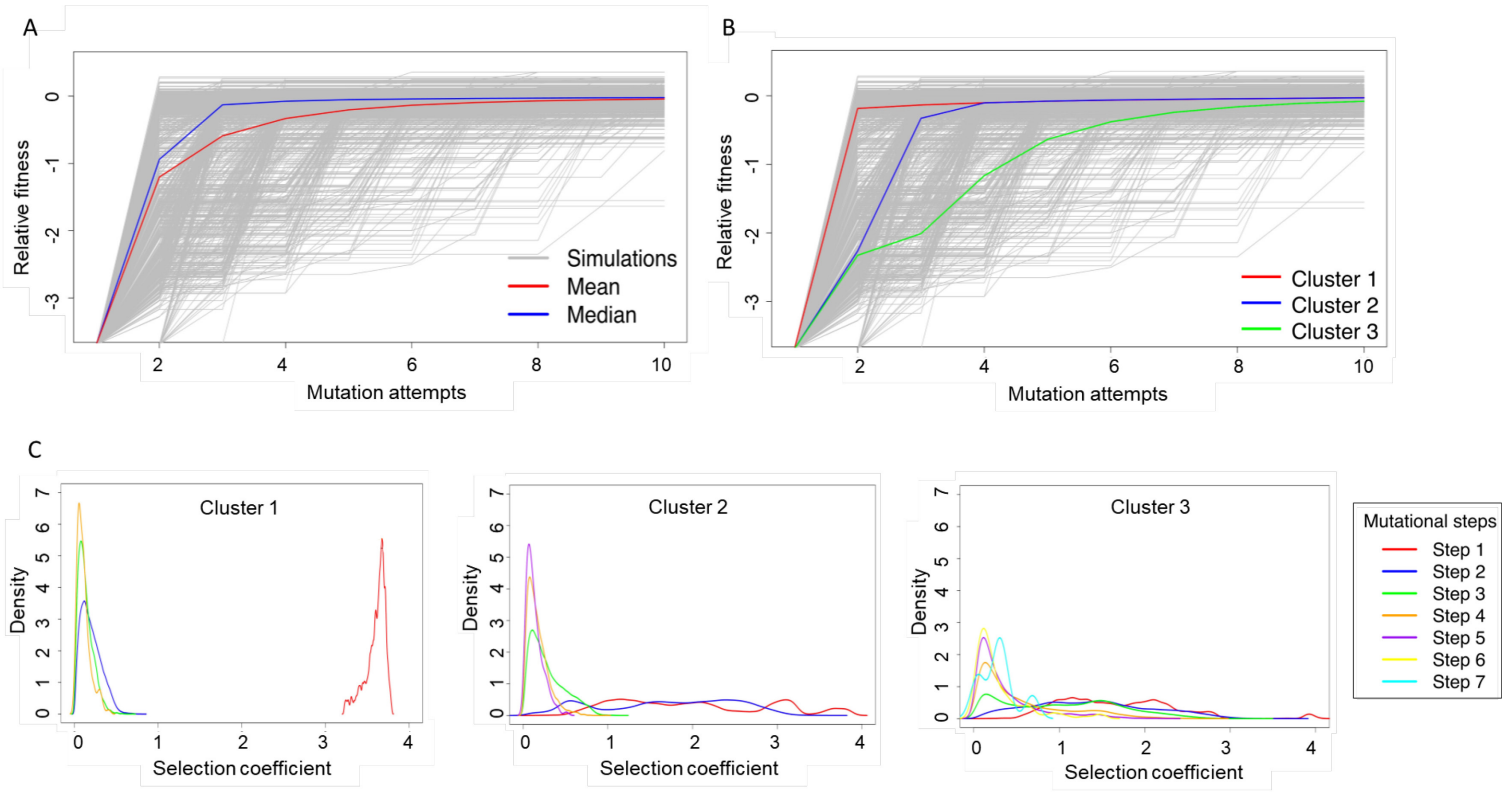

Figure 8. Trajectories from the deterministic evolutionary model form three distinct clusters.
The population dynamics simulation was performed for 10^6 trajectories. Each trajectory could go through 1000 mutation attempts. Shown are the first 10 mutation attempts from each trajectory. The mean (**A. red**) and median (**A. blue**) of the evolutionary trajectories are not sufficient to recapitulate the adaptive behavior of the population dynamics simulations. The evolution trajectories behave in a clustering trend at high concentration of Amp (2500 µg/mL) and high population size (10^6). The most populated cluster (**B. red**) includes 43% of the trajectories in which 1 mutation attempt is sufficient to confer resistance during adaptation. 29% of trajectories (**B. blue**) require at least 2 mutation attempts, and 28% require (**B. green**) 3 attempts or more. The density of selection coefficients at each mutation attempts was graphed, with each color representing a different mutation step. **C.** Cluster 1, the fast growth is dominated by a large selection coefficient at the first mutation and lower selection coefficients for subsequent steps. For the other clusters, there is a slow crawl to resistance. The specific pathway to resistance is dictated by the selection coefficient of their mutations.

Comparison with clinical isolates and with laboratory evolution

As each evolution trajectory is defined by a sequence of fixed mutations, we seek to determine which mutations are selected more frequently. Beneficial mutations that confer resistance should occur at high frequency. Therefore, we determine the correlation between mutations at high occurrence in the fast adaptation cluster, and mutations observed in clinical isolates and laboratory evolution. We want to determine if comprehensive fitness landscapes combined with our model can capture the resistance mutations reported in the literature.

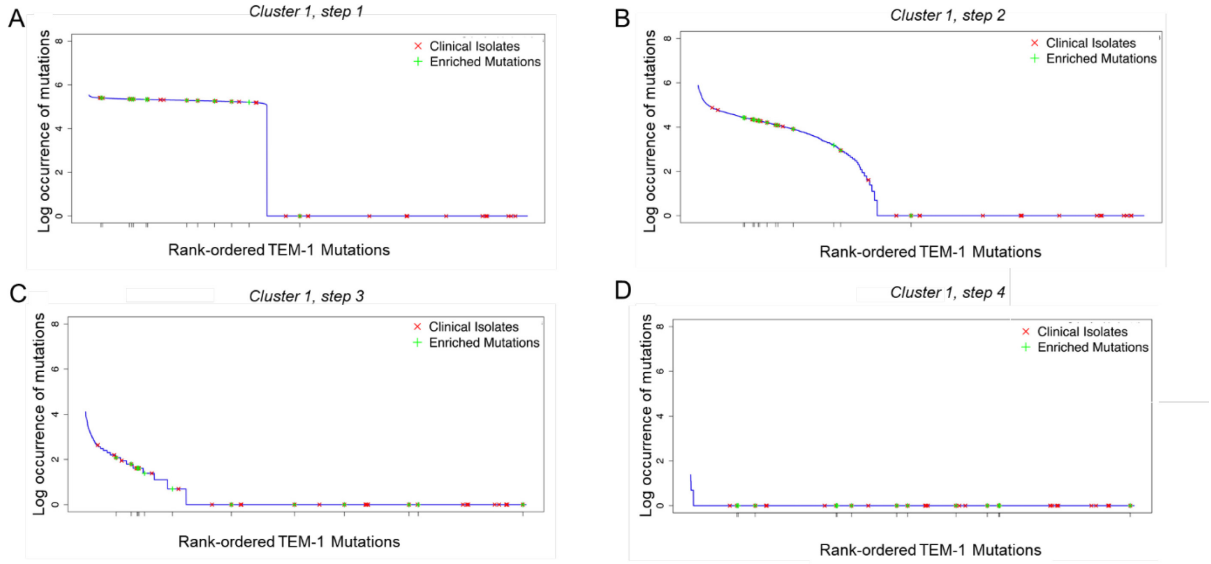

Figure 9. Clinical isolates and enriched mutations observed in the fast adaptation cluster.
The natural log occurrence of each mutation during the population dynamics simulations is graphed with the clinical isolates (**red**) and enriched mutations (**green**). We rank-ordered the TEM-1 mutations by their occurrence in the simulations. **A.** For cluster 1, at its first step, the enriched mutations are seen in high occurrence. **B.** For the second step, if resistance was not reached within the first step, the resistance mutation occurs at the second step. As adaptation progresses, the pool of available mutations that confer resistance diminishes as fitness is already at a maximum plateau. Ergo, step 3 (**C**), and 4 (**D**) show the low occurrence of clinical isolates and enriched mutations as they already occurred in step 1 and 2. We observe that mutants with a high occurrence count from the simulations are not necessarily conferring resistance. Currently, we cannot recapitulate mutations observed in clinical isolates.

For the first step, we observe that the occurrence rate per mutation is distributed into two segments: high occurrence and low or no occurrence, as shown in **Figure 9A**. The separation is due to the density of the selection coefficients. As only a few mutations can confer high fitness, the same mutations are selected at the first step of the simulations. We compare the high occurrence mutations against mutations reported from directed evolution and clinical isolates. The p-values reported from the two-sided Kolmogorov-Smirnov test (KS test) showed a significant relationship between high occurrence mutations at the first step and laboratory evolution mutations (p = 0,001) (Bershtein et al., 2008). We observed no relationship with clinical isolates (p = 0,4).

In the second step, fewer mutations are fixed at high occurrence, since the pool of available mutations is diminished (**Figure 9B**). The enriched mutations are still observed at a higher rate than the clinical isolates. There is a significant relationship between enriched mutations and high occurrence mutations (p=0,0003), and, to a lesser extent, between clinical isolates and high occurrence mutations (p=0,05). In the third step, we observe that enriched mutations and clinical isolates occur at a lower frequency, and are more disparate, as shown in **Figure 9C**. We did not observe any significant relationship between enriched mutations and high occurrence mutations (p = 0,2), and neither for clinical isolates (p = 0,5). For the last step in the fast adaptation cluster, only a few mutations are available mutations as the pool of potential mutations has run out. This is reflected in the low occurrence rate of mutations, as shown in **Figure 9D**. The enriched mutations and the clinical isolates are not reported in any of the mutations, and there is no relationship for either case (p = 1).

P-values calculated from two-sided KS-test in cluster 1

	Step 1	Step 2	Step 3	Step 4
Enriched	**$1{,}2 \times 10^{-3}$**	**$3{,}0 \times 10^{-4}$**	$2{,}0 \times 10^{-1}$	1
Clinical	$4{,}0 \times 10^{-1}$	**$4{,}7 \times 10^{-2}$**	$4{,}8 \times 10^{-1}$	1

Table II. The p-values for enriched mutations and clinical isolates were calculated with a two-sided KS-test in the first cluster.

We use the test to determine whether the samples were drawn from the same distribution. We set the rank-ordered occurrence rate of mutations in the simulations as the reference probability distribution. The occurrence rates of enriched mutations observed from laboratory evolution (Bershtein et al., 2008), and clinical isolates (Salverda et al., 2010) are assigned as sample sets, and we determine if there is a correlation between the reference and sample datasets. The p-values returned in step 1 and step 2 are significant for the enriched mutations. From the table, it is possible to determine that the enriched mutations observed in high frequency at step 1, and 2 are not random. Mutations from clinical isolates from ampicillin and all other selection agents observed were not recapitulated in any of the high occurrence mutations, except for step 2 of ampicillin. From step 4, the mutations observed are due to random chance. The purifying selection simulations are more accurate to recapitulate the enriched mutations.

Discussion

In this chapter, we seek to determine the potential pathways for the emergence of resistance by relating the biochemical effects of mutation on target genes, population demography, and single drug dosage. Our model shows that there is possibly pervasive epistasis in clinical isolates and mutants from laboratory evolution as we cannot fully recapitulate them in our single-step simulations. If there was no epistasis involved in clinical isolates mutants from laboratory evolution, our model should be able to recapitulate all first mutations arising in resistance. Therefore, we hypothesize that this behavior is due to the dependency of each mutation to its genetic background. Thus, simple and additive models of fitness such as our deterministic model cannot fully capture the evolutionary pathways as they do not consider the underlying epistatic networks involved in evolution. Nonetheless, in the context of purifying selection, comprehensive fitness landscapes capture a large fraction of enriched mutations, but only for selection in short-term evolution (Gupta and Adami, 2016). To summarize, we

develop an evolutionary model based on additive fitness to determine the pathways to drug resistance. We construct a comprehensive fitness landscape of TEM-1 under Amp selection from published DMS. By starting the simulations with a protein variant with low fitness, we allow competition between all mutants and perform adaptive walks on the fitness landscapes to study purifying selection and resistance.

Evolutionary trajectories follow three distinct growth rates

We constructed 10^6 evolutionary trajectories under high Amp concentration (2500 μg/mL) and at large population size ($N_e = 10^6$). Each trajectory represents a specific sequence of fixed mutations to reach resistance. The average and median of all evolutionary pathways do not reflect the evolutionary dynamics and dispersion of the pathways as they do not account for different growth rates. Instead, the pathways form three distinct clusters with different behaviors: fast adaptation, intermediate adaptation, and slow adaptation. The three clusters highlight the uniqueness of each pathway during purifying selection. The clustering behavior should be expected as proteins sharing mutations with common biochemical traits are shown to evolve parallelly, and often with the repeated acquisition of the same mutations (Harms and Thornton, 2013).

The fast adaptation cluster is populated by pathways which require in average only one mutation attempt to reach resistance. These pathways select and fix “blockbuster mutations” that confer high fitness gains. Blockbuster mutations or first-step mutations (the early mutations occurring in an evolutionary trajectory) are key mutations in an evolutionary pathway. They have been shown to cause extensive changes in gene expression and significantly impact the landscape’s shape for subsequent mutations (Rodriguez-Verdugo et al., 2015). These pathways contain first steps which are dominated by mutations with large selection coefficients, with subsequent steps contributing minimal fitness gains. This behavior is not observed in the other two clusters. The intermediate adaptation cluster is characterized by a climb to resistance, requiring in average two mutational attempts as it lacks “blockbuster mutations” in the first step, but is present in the second step. In average, the initial two steps work in complementation to reach resistance. The slow adaptation cluster has a different

behavior. None of the pathways contain a "blockbuster mutation" conferring significant fitness gains. Instead, they crawl slowly to resistance as their mutations do not have high selection coefficients. These simulations are accurate for evolution where genetic variation within a population is primarily supplied by mutations. Consequently, the model replicates mutation-driven evolution (Ueda et al., 2017).

Blockbuster mutations dictate the trajectories' behavior

As our simulations begin from a TEM-1 variant with low fitness, any mutation fixed initially will be beneficial and increase the overall fitness of TEM-1. The model will always favor beneficial mutations over deleterious mutations. Thus, reaching resistance is much faster with our current model, as only a few mutational attempts are required in most pathways. The increase in fitness diminishes the pool of available mutations since mutations conferring a lower fitness will not fix in the current model. This behavior is explained by the dynamic mutation-selection balance concept: an excess of strongly beneficial mutations will push a population towards higher fitness, and simultaneously, the available pool of beneficial mutation is depleted by adaptation (Goyal et al., 2012). Therefore, after the "blockbuster mutation", the number of available mutations that can increase the overall fitness of TEM-1 in Amp selection is much smaller. These initial mutations dictate the remaining available mutations for subsequent selection (Rodriguez-Verdugo et al., 2015; Harms and Thornton, 2013).

Clinical isolates are not recapitulated in the additive fitness model

We analyze the sequence of mutations of the fast adaptation cluster to study the effects of "blockbuster mutations" occurring in the first evolutionary steps. We compare the high occurrence mutations from the fast adaptation cluster to the resistance mutations observed in clinical isolates for Amp selection. We observe no statistically significant relationship between clinical isolates and high occurrence mutations. Due to pervasive epistasis, our evolutionary model cannot efficiently recapitulate the mutations observed in clinical isolates under Amp selection. The current evolutionary model and DMS do not account for the background on

which a mutation came from. It assumes the same genetic background for all beneficial or deleterious mutations, but Mackay (2014) has demonstrated that including epistatic constraints could greatly increase an evolutionary model's accuracy. Also, some resistance-conferring mutations are dependent on other mutations to fix in a population. For example, it was shown that mutations E104K, M182T and G238S in TEM-1 would not be considered strongly beneficial individually. However, when mutation E104K emerges conjointly in the population with mutations M182T and G238S, this TEM-1 derivative is significantly resistant to cefotaxime compared to the WT strain (Salverda et al., 2017). The model does not recapitulate such effects. Also, these results could be explained in terms of differences in experimental settings. A resistance mutation could reduce survivability in a real-world setting (clinical isolates), but not necessarily in a laboratory evolution experiment (Maharjan and Ferenci, 2017). These variabilities are important caveats to consider in the model. Thus, using the deterministic model, we could not identify mutations from clinical isolates under Amp selection.

Enriched mutations from directed evolution are recapitulated in the additive fitness model

In the context of enriched mutations in directed evolution, comprehensive fitness landscapes capture a large fraction of the adaptive effects. Consequently, we observe a strong relationship between high occurrence mutations in the fast adaptation cluster and enriched mutations in the first two mutational steps ($p < 0.001$). Subsequently, there was no relationship in the third and fourth steps ($p > 0.05$). Therefore, the highly occurrent "blockbuster mutations" in our simulations recapitulate the enriched mutations from directed evolution (Bershtein et al., 2008).

There are two possible explanations for the correlation between enriched mutations and the "blockbuster mutations" in the first two steps. The "blockbuster mutations" are significant due to their order in the evolutionary pathways **(**Rodriguez-Verdugo et al., 2015; Shah et al., 2015). If the first step is an enriched mutation, there is a lower chance of fixing another enriched mutation in the second step. In the other scenario, if an enriched mutation is not fixed

in the first step, it will be fixed in the second step. Therefore, with our current evolutionary model of additive fitness, we can determine the first passage time of drug resistance in a laboratory setting by identifying the first mutations arising under purifying selection.

Summary of the deterministic evolutionary model

To summarize, we observe pervasive epistasis in mutations from clinical isolates, and our model of additive fitness cannot fully capture these pathways due to these molecular biophysical constraints. Nonetheless, in the context of purifying selection, comprehensive fitness landscapes built from DMS capture a large fraction the mutations that are enriched for resistance. Using our evolutionary model, there is a possibility to predict which mutations would arise first under purifying selection. Identifying the first fixed mutation is primordial to set the background for subsequent mutations (Shah et al., 2015). Currently, our model is only accurate for purifying selection of single amino-acid mutagenesis. Due to the prevalence of epistasis during adaptation, it will be necessary to extend our model beyond single point mutations. With a complete evolutionary model and fitness landscape, we could potentially recapitulate the pathways observed in clinical isolates.

Chapter 4: Adaptive selection stochastic evolution model

From the previous chapter, we acknowledge the importance of the first fixed mutation in the evolutionary pathways to resistance. In the initial deterministic model, we only included the resistance level of TEM-1 against Amp and analyzed the emergence of resistance as a pathway. Thus, we could not accurately identify the clinical isolates mutations. In this chapter, we focus specifically on identifying resistance-conferring single point mutations. We update our initial model by including fitness cost and implement the Moran process. Here, we define fitness cost as the fitness of the bacteria with a mutation in the absence of antibiotics. The mapping of resistance level against fitness cost presented in **Figure 10**, can provide valuable insight into their inter-dependence on the emergence of resistance. Therefore, we include fitness cost in our new model to determine its role in the emergence of resistance but to study resistance-conferring single point mutations while in-treatment. Using the stochastic model, we found that fitness cost could play a role in alleviating the emergence of resistance. Although the model could not accurately identify all resistance-conferring mutations, it provides valuable insight into the development of treatments against the emergence of resistance.

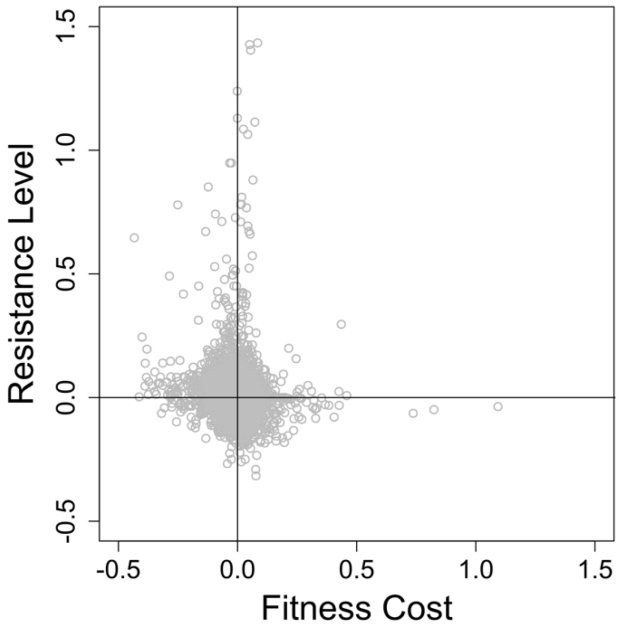

Figure 10. Resistance level against fitness cost mapping.
The x-axis corresponds to fitness effects in an environment without antibiotics while the y-axis corresponds to fitness effects in a cfx selection environment. Further studying the mapping will provide insight into the relationship between resistance level and fitness cost in the emergence of resistance. Fitness cost and resistance level values were obtained from Stiffler et al. (2015).

Methodology

We utilize the Moran process in our stochastic evolutionary model for its ability to simulate neutral drift in a finite population. The Moran model requires a fixed population of size N composed of two types of individuals: A and B, where A could represent WT strains, and B could represent resistance-conferring mutants. Both A and B reproduce at the same rate. At a specific time step, two individuals are selected randomly, one for reproduction and one for elimination, as shown in **Figure 11**.

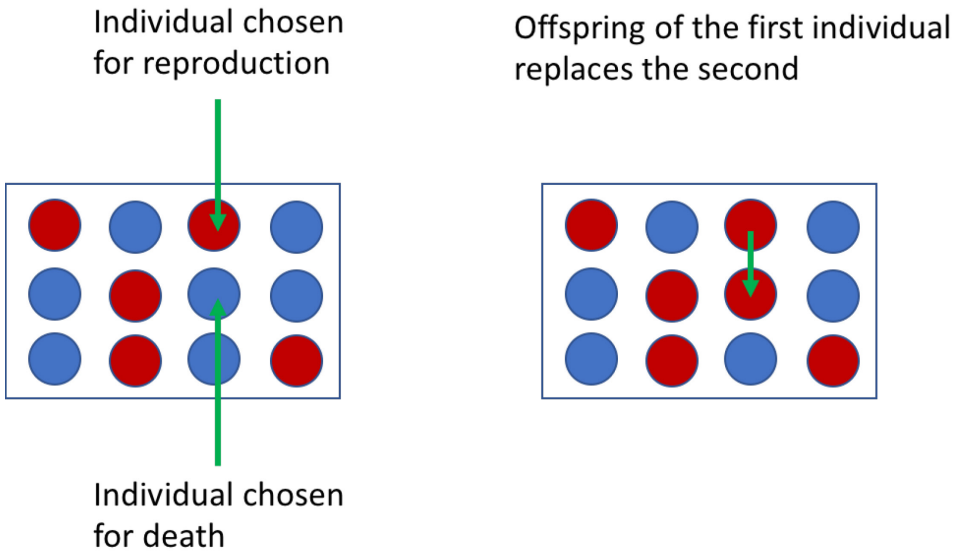

Figure 11. Moran model: a stochastic evolutionary model.
The Moran model represents the birth-death process to study selection in a finite population. In the first step, one individual is chosen for reproduction, and one is chosen for death. The individual selected for reproduction replaces the individual selected for death, as such, the population size remains constant throughout (Adapted from Nowak, 2006).

The constant birth-death process at each time step ensures that the population size remains constant. Therefore, the only stochastic variable in the model is the number of individuals A, denoted by i, and the number of individuals B denoted by $N - i$.

At any given time step, there are four possible scenarios that could happen:

1. An individual A is selected for both birth and death. The variable i remains the same in this case as the number of A individuals does not change. This event has a probability $(i/N)^2$ of occurring since each A individual in the population has a $1/N$ chance of being selected.
2. An individual B is selected for both birth and death. The variable i also remains the same in this case as the number of B individuals does not change. This event has a probability $[(N\text{-}i)/N]^2$ of occurring since each B individual would have an $(N\text{-}i)/N$ chance of being selected.

3. An individual A is chosen for birth, and an individual B is chosen for death. The variable i has changed to $i + 1$ as there is one more A individual in the population. This event has a probability $i(N - i)/N^2$ of occurring.
4. An individual A is chosen for death, and an individual B is chosen for birth. The variable i has changed to $i + 1$ as there is one less A individual in the population. This event has a probability $i(N - i)/N^2$ of occurring.

The stochastic matrix, $P = [p_{ij}]$, of size $(N + 1)$ x $(N + 1)$, represents the probabilities of moving from the state i to any other of the three states: i, $i + 1$, or $i - 1$. The previously described scenarios determine the four possible transitions in the matrix. The defining property of the birth-death processes is that in any of the step, the state variable i can only change by one until it reaches either absorbing states.

There are two absorbing states in the Moran model, defined by $i = 0$, and $i = N$. The transient states are defined by $i = 1, \ldots, N - 1$. The birth-death process remains in the transient states until it reaches an absorbing state. Once the process reaches either of the two absorbing states, it will remain there indefinitely. The birth-death process of the Moran model is presented in **Figure 12**. Given enough time, the whole population will consist of either individual A or B as coexistence is not possible (Nowak, 2006). Therefore, the Moran model can provide valuable insight into the mechanisms of the emergence of resistance as it can replicate the competition between a resistance-conferring mutation and wild-type strains.

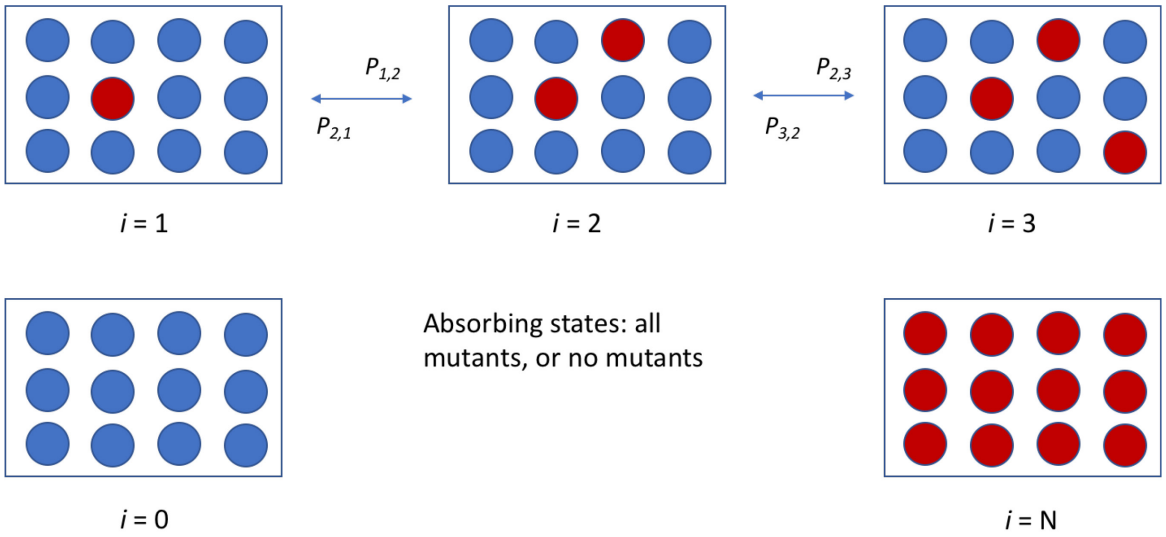

Figure 12. The Moran model consists of a birth-death process.
The Moran model is a birth-death process. A mutant can change by one at most, thus the state variable (i) can either remain unchanged or move to i-1 or i+1. There are two absorbing states, either $i = 0$ and $i = N$. Whenever an individual reaches an absorbing state, further changes in the population cannot occur unless a mutation is included. (Adapted from Nowak, 2006).

Tau-leaping to solve stochastic evolutionary models

The Moran process is used as a stochastic model to study the evolutionary dynamics of a population at a constant size. It can perform numerous neutral evolutionary steps before any significant changes in the population are observed. In probability theory, tau-leaping (τ-leaping) is an approximate method to efficiently simulate large stochastic systems (Gillespie, 2001). The basic tau-leaping method as described by Gillespie consists of three main steps:

1. Select a value for τ, the time step, that satisfies the Leap Condition. Satisfying the leap condition consists of identifying a temporal leap by τ which will result in a state change λ which is effectively infinitesimal.
2. Calculate the Poisson approximation to determine the net change in state λ, which is the number of times each event occurs during the determined interval.
3. Complete the leap by replacing time t with $t + \tau$, and x by $x + \lambda$, where x is the current state.

It is important to satisfy the leap condition with a tau that allows many events to occur. This results in a leap in the system, instead of a single step-event (Gillespie, 2001). By using a Poisson approximation, the tau-leaping method can leap over many fast events and approximate the stochastic behavior of the system (Cao et al., 2007). The tau-leaping method is used to simulate the numerous neutral steps in the Moran process. This significantly accelerates the simulations as we only analyze the time points where relevant changes in the population occurred (Cao et al., 2007). Thus, we use the adaptive tau-leaping solver to approximate the evolutionary trajectories from the Moran model.

Tau-Leaping Stochastic Solver

To solve the Moran process with tau-leaping, we used the R package, Adaptivetau – Tau-Leaping Stochastic Simulation (Johnson, 2016). The adaptivetau package implements both an exact solution and an approximate solution known as the "adaptive tau-leaping algorithm" to simulate the Moran process. We employed the approximate solution in our simulations to increase simulation speed while maintaining reasonable accuracy.

Stochastic evolutionary model algorithm for adaptive selection

For our stochastic evolutionary model, we set A to the number of mutants and B to the numbers of WT strains. The model returns the fraction of mutants in a population as a function of the total population size (A/N). Our algorithm requires four different inputs: N, f_0, s, and t, where N is the effective population size, f_0 is the fraction of mutants composing the population, s is the selection coefficient, and t is the total amount of time in hours for which the tau-leaping should be performed. The algorithm determines the drift rate for fixation of a mutant. Then, it returns the fraction of the population that is mutant after t hours.

For the total amount of t hours required, we perform the adaptive tau-leaping 100 times for every hour, and the average fraction of mutants is recorded for each hour. The algorithm for the stochastic evolutionary model is presented in **Figure 13**.

As we are modeling the Moran process, we set two critical values as the absorbing values: $X_1 = 1/N$ for the lower boundary, and $X_2 = 1-1/N$ for the upper boundary. Therefore, we implement three cases when calculating the fraction of mutants in a population. The first case is for fractions below the lower boundary; the second case is for fractions above the upper boundary, and the third case is for fractions within the boundaries. For the first case, if the reported fraction is lower than X_1, we move the population near the lower boundary with a right-continuous transition matrix of +1. For the second case, if the fraction is higher than X_2, we move the population near the upper boundary condition with a left-continuous matrix of -1. When the fraction of the population is within the conditions of the third case, we perform the Moran process with a transition matrix of -1 and +1.

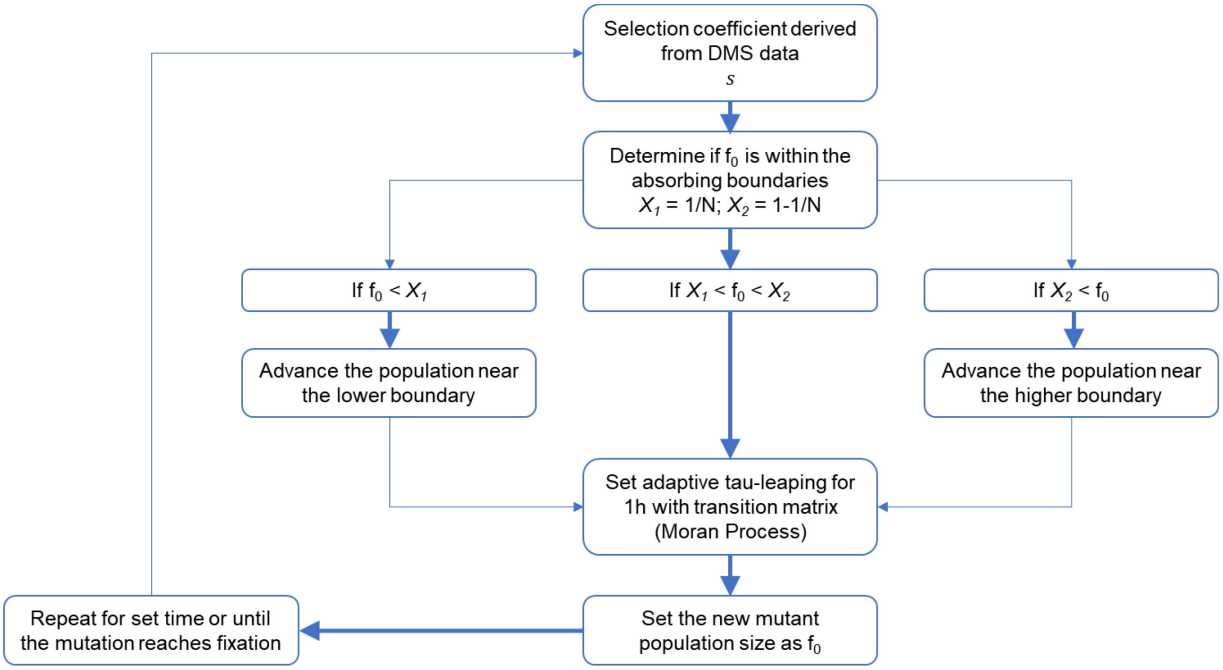

Figure 13. Algorithm for the stochastic evolutionary model.
We improve on the initial model of microbial evolution by implementing the stochastic evolutionary process of the Moran model. We simulate the growth of each TEM-1 mutants by allowing competition with WT strains only. We use the adaptive tau-leaping stochastic solver to replicate the evolutionary process of the Moran model and to accelerate simulation speeds.

Fitness landscape of TEM-1 β-lactamase under cefotaxime selection

The algorithm for the Moran process was written in R 3.3.3 (R Core Team, 2013). The model requires two fitness landscapes: one with the antibiotic selection and one without selection. The two landscapes are referred to as resistance level (RL) and fitness cost (FC), respectively. The antibiotic used in the resistance level is cefotaxime (cfx). Cefotaxime is a β-lactam antibiotic and is part of the third-generation class of cephalosporins. It is often used to treat penicillin and penicillin-derivate resistance (Ma et al., 2002). In this case, cfx is used, as TEM-1 does not have resistance against this new antibiotic. Thus, we can study how TEM-1 will gain resistance specifically against it. The DMS at 0,2 μg/ml cfx selection is used to represent

RL. The DMS without antibiotic selection (0 µg/ml cfx) is used to represent FC. Both datasets are obtained from Stiffler et al., (2015) and are presented in **Figure 14**.

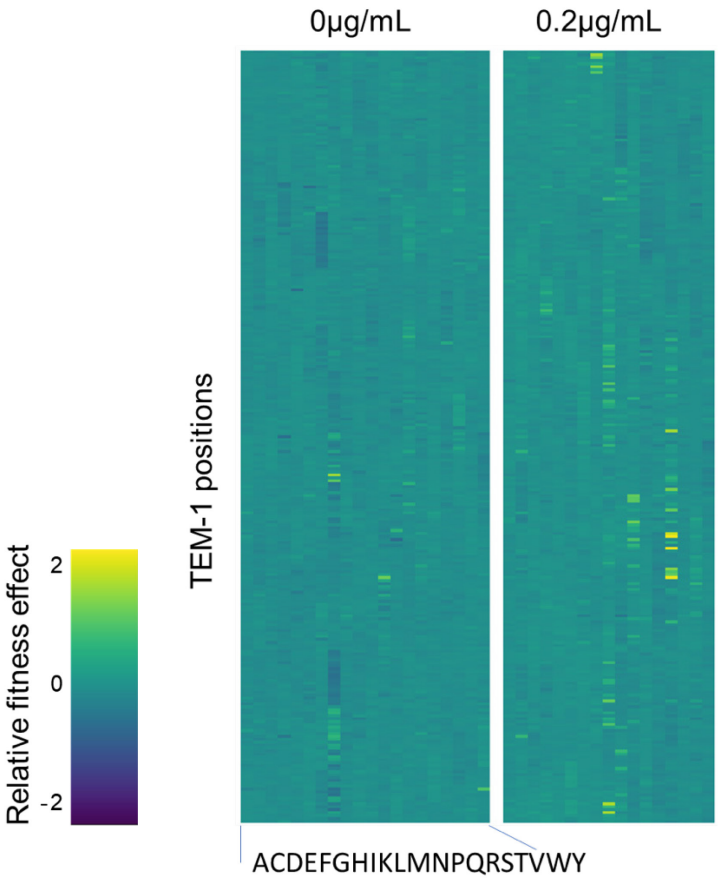

Figure 14. Comprehensive fitness landscape of TEM-1 under cefotaxime selection. Each row represents a TEM-1 position, and each column is one of the twenty possible amino acid mutation. Beneficial mutations are highlighted in yellow while deleterious mutations are highlighted in blue. The fitness landscape is used in our stochastic evolutionary model. For our simulations, the 0 µg/mL is used as the fitness cost dataset, while the 0,2 µg/mL is used as the resistance level dataset. Both datasets are represented as relative fitness effects, calculated with Eq. 1. (Adapted from Stiffler et al., 2015).

Adaptive evolution simulations

Using the Moran model, we want to determine which TEM-1 mutants would outcompete the WT strain in an oscillating environment and adapt to possibly confer resistance. Thus, we allowed competition only between the WT strain and each individual TEM-1 mutant and

determined if the mutant could propagate in the population. We determined the composition of each of the 4997 TEM-1 mutants in a population by simulating its growth. The effective population size N is set to 10^7. Each mutant is assigned a starting fraction of $1/N$. We used an oscillating treatment to simulate the effects of antibiotics during treatment. Each mutant is simulated in an oscillating treatment for 48 h. For the first 24 h, we use the fitness effects from the RL dataset, and then for the next 24 h, we use the fitness effects from the FC dataset. The final fraction of mutants is recorded after 48 h. As we wished to simulate the dose application of antibiotics during treatment, we converted the number of generations to hours. The script for the simulations was written in Python, and simulations were performed on an iMac, version 10.12.6 with 3,2 GHz Intel Core i5, and 16GB memory. The algorithm outputs a matrix where each row is the growth behavior of a mutant and each column represents the time. Each cell in the matrix is the fraction of mutant observed in the population.

Evolution trajectory analysis

To determine if the pathways were grouped depending on their growth rates, all evolutionary pathways are clustered with the kmlShape R package (Genolini et al, 2016). The package clusters longitudinal data using shape-respecting distance. The kmlShape algorithm was set to generate the optimal number of clusters with the designated sample of trajectories. To optimize clustering run times, the sample size is set to a minimum of 2500 trajectories as calculated with a confidence level of 95%. The centroid of each cluster is used as a reference to assign the complete set of 4997 evolutionary trajectories.

Model analysis

Model analysis is performed in R and Python. We calculated a mutant's probability of survival through its fraction in the population. We stipulated that the higher the fraction, the higher the probability for the mutant to fix in a population. When a mutant composes at least 50% of its population, the mutant has a higher probability of surviving over the WT strain. If the mutant reaches the 50% threshold, we assumed that it will eventually take over its population. This threshold is referred to as the fixation threshold in our model. Although a more stringent

criterion of 95%, similar to the one applied by Cowperthwaite et al. (2006) could be used to determine fixed mutations in our simulations, it would be impractical as we are simulating competition between the WT strain and one mutant strain only. These simulations are only intended to serve as an approximation of which TEM-1 mutants could provide resistance against cfx. Therefore, when a TEM-1 mutant overtook a population and consisted of at least 50% of the population at the end of the simulations, we assumed that the mutation would eventually fix if given enough time. Although this is not always the case due to the stochastic behavior of evolution and genetic drift, we assumed this low fixation threshold to include as many mutants as possible to supplement our analysis. Finally, the clinical isolates and laboratory evolution mutants under cfx selection are obtained from Salverda et al. (2010) and are presented in **Table 3.** These mutants are compared against the mutants from our simulations that reach above the fixation threshold for a higher probability of survival.

Clinical isolates	Clinical Isolates (Cont)	Clinical Isolates (Cont)	Lab isolates	Lab Isolates (Cont)	Lab Isolates (Cont)
E28D	D163G	R244G	D35A	S124R	M182L
E28K	D163H	R244H	D38G	S130G	M182R
K34E	R164C	R244L	D38N	P145L	M182T
D35P	R164H	R244S	D38V	P145Q	M182V
D38N	R164S	V262I	D38Y	H153D	G196S
Q39K	T265M	S268G	Q39K	H153L	A224V
A42V	W165C	R275A	Q39R	H153Q	A237S
L49M	W165R	R275L	A42G	H153R	A237T
L51P	I173V	R275Q	L49M	H153Y	G238A
E64K	N175I	N276D	L49V	H158Y	G238S
E64K	D179E	N276R	L51F	D163V	E240G
M69I	M182T	N276S	L51I	R164C	E240K
M69L	G196D	A280V	M69L	R164G	E240V
M69V	G196S	A284G	V80I	R164H	R244C
G92D	R204Q		G92D	R164N	R244S
N100S	G218E		G92S	R164S	T265A
L102V	L221M		N100G	R164Y	T265M
E104K	A224V		N100K	W165G	S268G
T114P	A237G		L102M	I173L	S268N
D115G	A237T		E104A	I173R	S268T
S124N	G238D		E104G	I173T	R275L
I127V	G238N		E104K	I173V	R275Q
S130G	G238S		E104V	N175D	N276D
P145A	E240K		T114A	N175Y	N276S
H153R	E240R		T114M	D179G	A280T
D157E	E240V		S124G	D179Y	
H158N	R244C		S124N	M182K	

Table III. Clinical isolates and laboratory evolution mutations for TEM-1 under Cfx selection.
The mutations were sampled from the literature on TEM-1 resistance (Salverda et al., 2010). The table is used to compare against our results from the simulations to determine if the stochastic model could accurately identify mutations observed in the literature. The mutations are presented in alphabetical order.

Oscillating treatment with different conditions

Using the stochastic evolutionary model, we tested multiple time constants to determine an optimal regimen against resistance. We simulated the evolutionary trajectories of the 4997

TEM-1 mutants for 48 h with different oscillating time constants: 4 h, 6 h, 8 h, and 12 h. The algorithm outputs the evolutionary trajectories and final fraction for each mutant depending on the time constant used.

To determine the role of fitness cost in the emergence of resistance, all simulations were repeated without FC. All FC values were set to null for the simulations. We performed the simulations for 48 h with the same population size, starting mutant fractions, and time constants.

Finally, we performed the simulations for a total of 120 h to generate the phase-space of resistance at a constant concentration with different time constants. The phase-space represents all the possible states of the simulations by combining all the time constants used in the model. The phase-space provided us with a better understanding of the dependence of time and fitness effects on the emergence of resistance.

Results

Adaptive selection under cefotaxime selection form clusters

In this chapter, we build a stochastic evolutionary model using the comprehensive fitness landscape of TEM-1 under cfx selection. Contrary to the first model of additive fitness, we simulate the evolutionary trajectories of single-site mutants alone. We only allow competition between a mutant strain and the WT strain. Competing each mutant against WT allows us to determine which mutants would emerge from a population and would possibly confer resistance. We do not allow competition between all the mutants. The trajectories of the mutants are simulated for a total of 48 h: 24 h in the resistance level regime, followed by 24 h in the fitness cost regime. Consequently, the evolutionary pathways of each mutant are dictated by the fitness effects of resistance level and fitness cost.

We observe that the trajectories form three distinct clusters as shown in **Figure 15A**. We identify three cluster centroids that recapitulate the pathways. The first cluster represents 95% of the pathways and correspond to weakly beneficial mutations and are characterized by poor growth in their population (**Figure 15A**, red). This group represents all the mutants without a significant impact on their population after 48 h. Their fitness effects in either regime are too weak. The second cluster represents 3% of the pathways and corresponds to the mutants which reached the fixation threshold within the first 24 h (**Figure 15A,** blue). The third cluster represents 2% of the pathways and corresponds to mutants which reached the fixation threshold after the change in environments (**Figure 15A**, green).

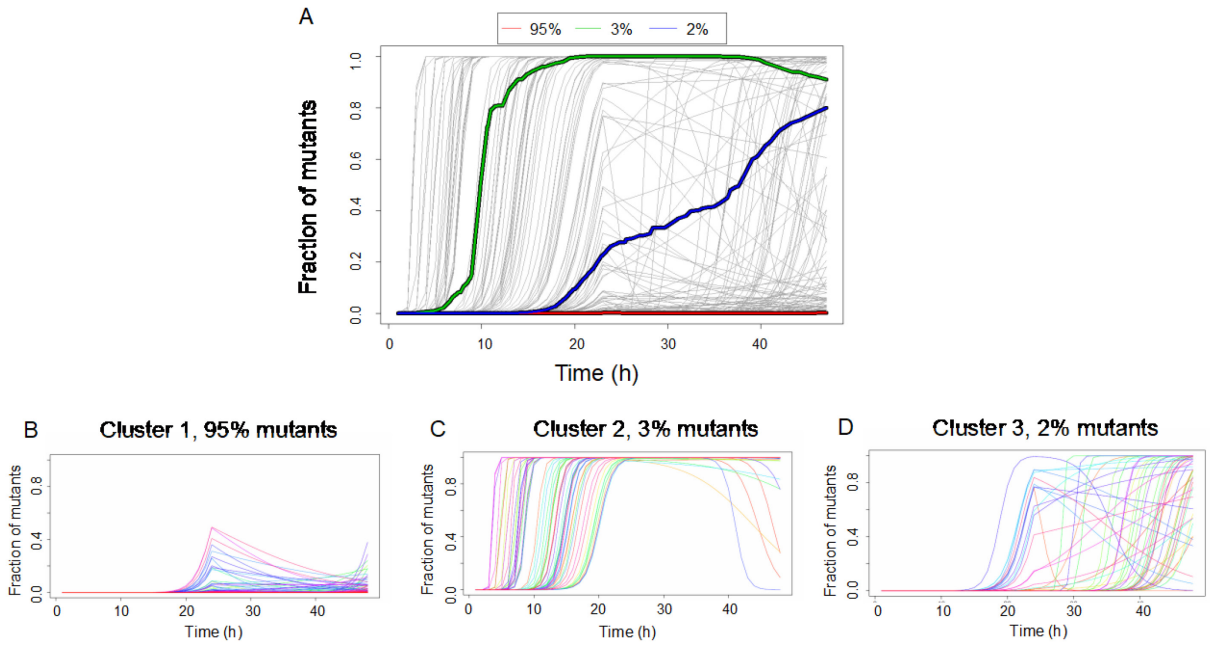

Figure 15. Trajectories from the stochastic evolutionary model form three distinct clusters.
We simulated the growth behavior of all TEM-1 mutants for a total of 48 h. The environment was switched from selection to selection-free after 24 h. We clustered the trajectories as a time series to group trajectories with similar growth behavior. X-axis: Time in hours of the simulations. Y-axis: Fraction of mutants composing a population. **A.** Three distinct clusters are formed in our simulations. The pathways are grouped depending on their growth rate using Frechet's clustering. **B.** The no growth cluster groups all mutants without significant fitness effects in either regime. **C.** This cluster groups all mutants with high fitness effects in the selection regime. The mutants reach the 50% threshold within the first 24 h hours. **D.** This cluster groups all mutants with high fitness effects on the fitness cost regime. The mutants reach the 50% threshold within 24 h following the change in environment.

From the initial simulations, we observe that the mutants are grouped by the behavior of their evolutionary pathways. We identify three groups: mutants with weak fitness effects in both regimes; mutants with strong fitness effects in the resistance level regime; and mutants with strong fitness effects on the fitness cost regime. In general, the mutants require only strong fitness effects in either resistance level or fitness cost to take over their population. The evolutionary trajectory of each mutant is determined by the fitness effects in either regime, as shown in the resistance level and fitness cost map.

Resistance level against fitness cost mapping

The mapping of resistance level against the fitness cost with the clusters is presented in **Figure 16**. The first cluster of mutants is centered around neutral fitness effects due to minimal growth in their population. The second group of mutants is characterized by high fitness in the resistance level and are grouped along the y-axis. We observe that the high resistance level allows for some mutants to offset a negative fitness effect in the fitness cost regime. The third cluster is characterized by mutants with high fitness effects in the fitness cost regime and is grouped along the x-axis. We observe that the high fitness effects on the fitness cost regime can counterbalance the low resistance level. We notice that a high fitness effect in one regime can offset a negative fitness effect in the other, consequently giving more chances for mutants to emerge in a population. Thus, the mapping of mutants is dependent on the fitness effects of either regime.

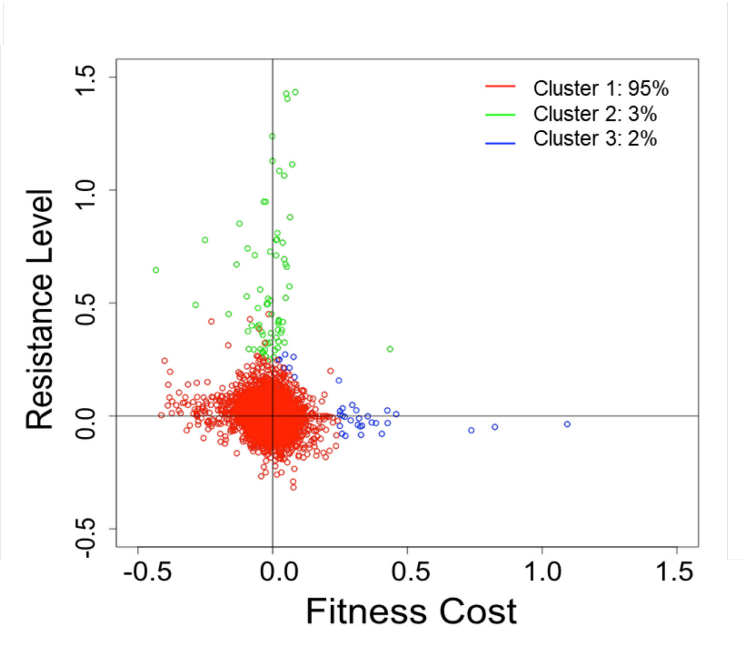

Figure 16. Survival probability in oscillating antibiotic concentration depends on the trade-off between resistance level and fitness cost.
Clustering the mutants on the resistance level against fitness cost mapping highlights the dependency of the mutants to the fitness effects of each regime. We can identify the relevant mutants of interest in the emergence of resistance whether we are interested in resistance level or fitness cost.

Performing evolutionary simulations with different time constants of drug administration

To identify optimal dosage regimens, we elucidate the role of the time constants in our evolutionary simulations. We determine the effects of different time constants on the evolutionary behavior of the mutants. We perform the simulations at four different time constants: 4 h, 6 h, 8 h, and 12 h. Each simulation is performed for a total of 48 h with an oscillation between resistance level and fitness cost at the specified time constant. The evolutionary trajectories of each mutant for the different oscillating tau are presented in **Figure 17**. We counted the number of mutants which reached the defined 50% threshold, at the 24 h and 48 h timepoint. The counts are presented in **Figure 18**.

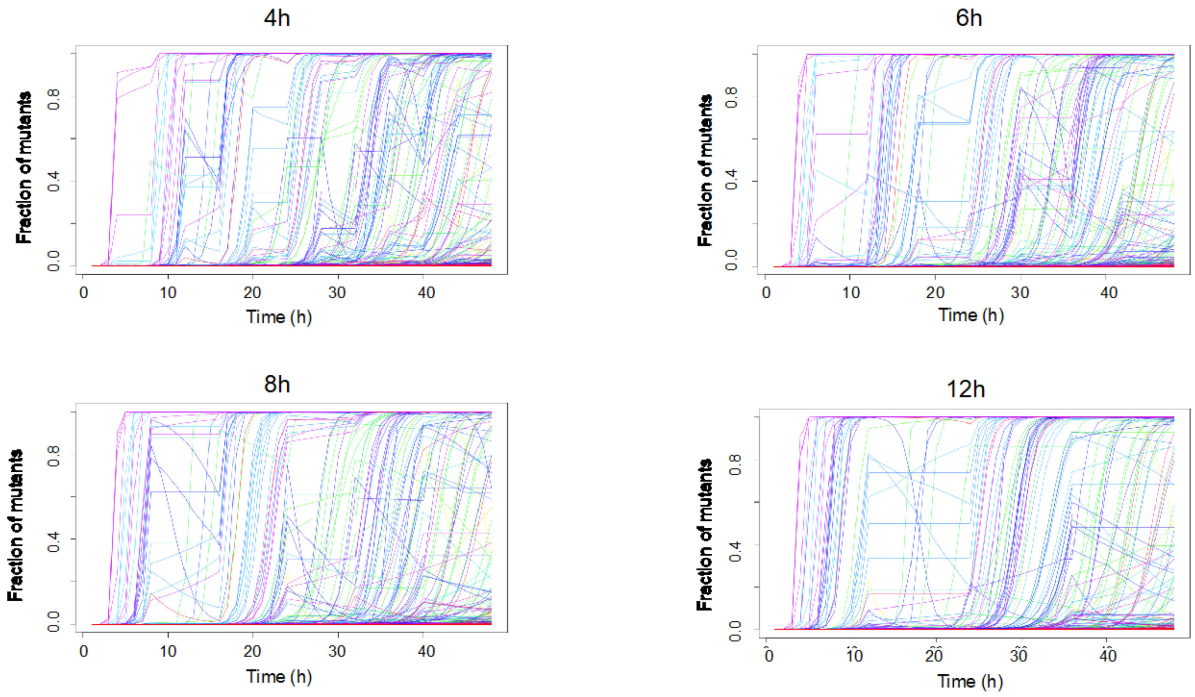

Figure 17. Evolutionary trajectories using different time constants.
We simulated the growth behavior of all TEM-1 mutants for a total of 48 h. The environment was switched from selection to selection-free at the specific time constant for each, either at 4 h, 6 h, 8 h, or 12 h. X-axis: Time in hours of the simulations. The environment was switched from selection to selection-free depending on their tau. Y-axis: Fraction of mutants composing a population. Different time constants confer different evolutionary behaviors to the emergence of resistance.

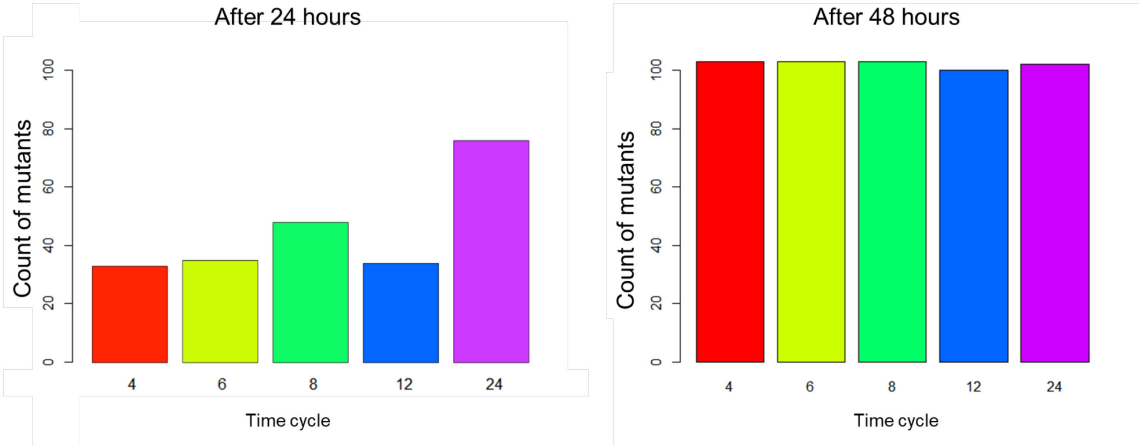

Figure 18. Count of mutants reaching the 50% threshold after 24 h and 48 h.
We count the number of mutants that reach the fixation threshold while competing against WT strains, and with a higher probability of survival, depending on the time constant used in the simulations. After 24 h, the counts between time constants seem all different, but given enough time, the different cycles reach a similar number of mutants. Stable oscillation at a specific time constant is not sufficient to purge resistant mutants from a population in our model.

After 24 h, the simulations with the 8 h tau captured the highest numbers of mutants with increased probability of survival, while the 12 h tau had the fewest number of mutants to reach the same threshold. We observe that the resistance level regime is mainly responsible for the increase of mutants in their population. After 48 h, regardless of the tau used, all the simulations reached a similar number of mutants with a higher probability of fixation at an average of 102,2. Therefore, in our model, using different time constants in the simulations was not sufficient to alleviate the emergence of resistance, or to reduce the overall number of mutants with a higher probability of fixation. Interestingly, at the end of the simulations, we observe a set of common mutants between the different tau. The current stochastic model captures four of the five most common clinical isolates mutants under cfx selection in the literature: E104K, R164H, R164S, and G238S.

Determining the role of fitness cost in the evolutionary model

We analyze the resistance level and the fitness cost of the clinical isolates to determine why the fifth most common clinical isolate, M182T, was not recapitulated in our model. We observe that the four clinical isolate mutants have high fitness effects in the resistance level, but negligible fitness cost effects (**Table 4**). This is not the case for M182T which has neutral fitness effects in both regimes. Therefore, we want to determine the role of fitness cost in resistance-conferring mutations.

Mutation	Resistance Level	Fitness Cost
E104K	0.5734668	0.06217737
R164H	1.064041	0.04283956
R164S	1.129395	-0.000719945
G238S	1.427253	0.0506182
M182T	-0.0270535	0.05449325

Table IV. Resistance level and fitness cost of the most common mutations observed in clinical isolates.
The most common mutations from clinical isolates (Salverda et al., 2010) have a high resistance level and a low fitness cost from the DMS datasets we are using. Mutation M182T is considered neutral compared to the other four mutations and thus does not emerge in our simulations.

From the simulations, we observe that clinical isolates are dependent on the resistance level regime only. We seek to determine the mutants' behavior if fitness cost was removed from the simulations. Thus, we set the fitness effects of mutations in the antibiotic-free regimen to null. The evolutionary trajectories without fitness cost are presented in **Figure 19A**. We observe a noticeable lack of growth or decline in the fitness cost regime due to the absence of selection. Also, the provided time was not large enough for the populations to drift to either absorbing states. Nonetheless, we obtain 4997 distinct evolutionary pathways from the simulations without fitness cost.

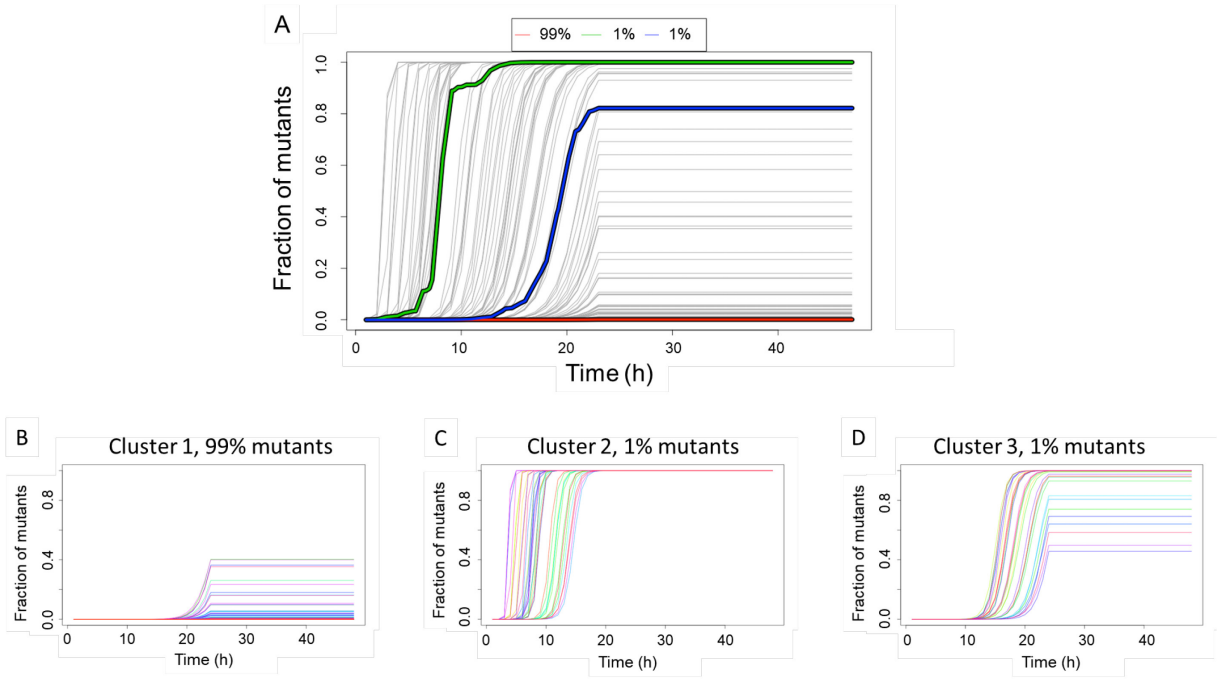

Figure 19. Evolutionary trajectories without fitness cost in the simulations. We simulated the growth behavior of all TEM-1 mutants for a total of 48 h. The environment was switched from selection to neutral drift after 24 h (Fitness cost values were set to null). Even without fitness cost, trajectories follow three different pathways. X-axis: Time in hours of the simulations. The environment was switched from selection to selection-free after 24 h. Y-axis: Fraction of mutants composing a population. **A.** Three distinct clusters are formed in our simulations. The pathways are grouped depending on their growth rate. **B.** The no growth cluster groups all mutants without significant fitness effects in either regime. **C.** This clusters groups all mutants with high fitness effects in the selection regime. **D.** This clusters groups all mutants with intermediate fitness effects in the resistance level regime.

We group the evolutionary pathways without fitness cost and identify three cluster centroids as shown in **Figure 19A**. We observe that the pathways form three distinct clusters. The first cluster represents 99% of the pathways and corresponds to the mutants that did not take over their populations after 48 h (**Figure 19B**). This group represents the great majority of mutants without a significant impact on their population. The second cluster represents 1% of the pathways and corresponds to mutants which fix in their population within the first 24 h of the

simulations (**Figure 19C**). The third cluster represents 1% of the pathways and corresponds to the mutants which reach the fixation threshold but at a slower rate (**Figure 19D**). These mutants are considered second-rate as they grow slower than the fast group. The slow growth is due to a lower average of fitness effects in the resistance level stage. The fast-growing group has a higher average of fitness effects in the resistance level compared to the slow-growing group. Thus, removing fitness cost affects the different growth behaviors of mutants. We then determine the difference in mutant counts that reached the fixation threshold between the simulations with FC and without FC.

We calculate the number of mutants with an increased probability of survival after 48 h of the oscillating regimen. We compare the counts from the simulations with and without fitness cost (**Figure 20**). For the simulations without fitness cost, the average number of mutants between all tau is 77,4 mutants with a variance of 9,3. For the simulations with fitness cost, the average number of mutants between all tau is 102,2 with a variance of 1,7. Thus, we observe a significant drop in overall mutant counts when fitness cost is not included in the simulations. Although the average number of mutants has changed, we seek to determine if the average fitness effects of mutants changed too.

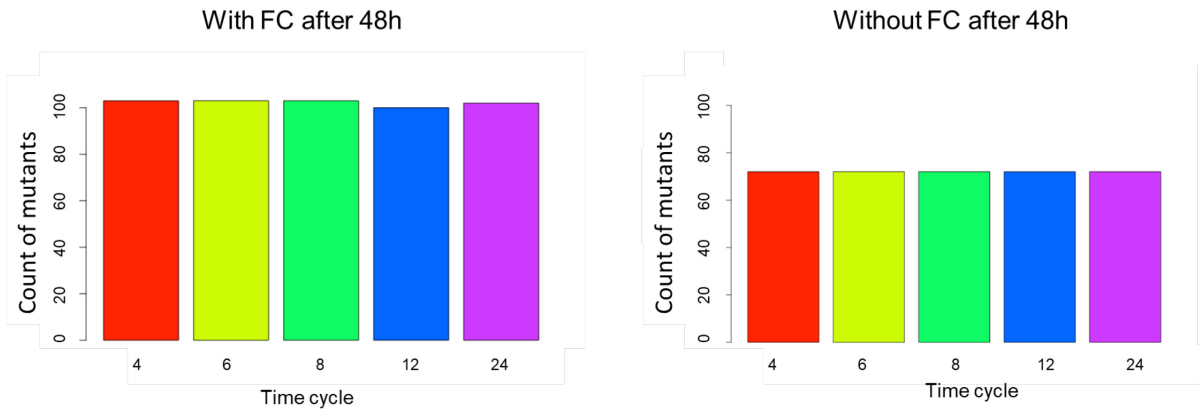

Figure 20. Mutant count following 48 h evolutionary simulations with fitness cost (FC) and without FC.
We count the number of mutants with a higher probability of survival, depending on their time constant, from our two simulations with FC and without FC. Fitness Cost increases the overall number of mutants. The simulations with FC have a higher average number of mutants than the simulations without FC. Fitness cost may play a role in maintaining resistant mutants in the population.

To determine the influence of fitness cost on the overall quality of the mutants, we analyze the average fitness effects of the ones reaching the fixation threshold, as shown in **Table 5**. Interestingly, although we record a higher number of mutants for the simulations with fitness cost, we observe a lower overall average of fitness effects in the resistance level, at 0,40. Compared to the simulations without fitness cost, we observe a higher average of fitness effects in the resistance level, at 0,55.

	After 48h with fitness cost		After 48h without fitness cost	
Time Constant	Resistance Level	Fitness Cost	Resistance Level	Fitness Cost
4	0.4115081	0.09283655	0.5597318	-0.02144223
6	0.4122767	0.09052067	0.5450198	-0.02661622
8	0.4048036	0.0972031	0.5482018	-0.02751765
12	0.4019312	0.1049945	0.5609108	-0.02842964
24	0.3889759	0.1095555	0.5589774	-0.02879229
Average	0.4038991	0.099022064	0.55456832	-0.026559606

Table V. Average fitness effects for the stochastic simulations to study the effects of fitness cost.
Simulations without fitness cost increased the average fitness effects of the resistance level, leading to mutants with higher resistance. Fitness cost reduces the average fitness effects seen in the resistance level regime, leading to an overall lower quality of resistance-conferring mutants as the pool of available mutants is bigger.

Survival probability of mutants and duration of antibiotic administration

To determine the comprehensive effects of an oscillating treatment on a population, we construct the phase-space of resistance of TEM-1 under cfx selection. We combine the different tau in our model to simulate the evolutionary trajectories of mutants for a total of 120 h. For example, the 4 h tau is combined with the 6 h and we performed the oscillating simulations for 120 h, with 4 h in the RL and 6 h in the FC. The 4 h tau is then combined with the 8 h tau, and we repeated the simulations. This process is performed for all the different tau. The phase-space provides us with important parameters to better understand the optimal combinations of tau to reduce the number of potentially resistance-conferring mutants reaching fixation. Each state in the phase-space corresponds to a specific combination of time constants. The phase-space of TEM-1 resistance to cfx is presented in **Figure 21A**, and we record the number of mutants which reached our defined fixation threshold for each state. To study the dependency of each state to the fitness effects of selection, we generate the mapping of the resistance level against the fitness cost of the mutants as shown in **Figure 21B**.

We observe that the tau in each regime dictates the final number of mutants with increased survivability. States combining a large tau at the resistance level with a smaller tau at the fitness cost, contain the highest number of mutants with an increased probability of survival. The higher number of mutations fixing increases the probability of having resistance-conferring mutants in the population. The highest number of mutants is recorded for the simulations oscillating between resistance level for 24 h, and fitness cost for 4 h. Concurrently, the lowest number of mutants recorded is for the simulations oscillating between resistance level for 6 h and fitness cost for 12 h. We observe that the number of mutants in each state decrease as we increase the tau for fitness cost and decrease the tau for resistance level. This shift in behavior is also observed in the mapping of fitness effects.

The states in the phase-space with the lowest number of mutants are characterized by high fitness effects in both the resistance level and fitness cost regimes. The high fitness effects in the fitness cost regime counteract the mutants with negative fitness effects in the resistance level regime in the simulations. Similarly, the high fitness effects in resistance level can neutralize the mutants with negative fitness effects on the fitness cost regime. As the tau shifts to higher fitness effects in resistance level regime and lower fitness effects in the fitness cost regime, the composition of each state is also shifted. The mutants that arise in these states are characterized by high fitness effects in the resistance level. This overwhelming positive effect for the resistance level enables mutants with negative fitness effects on the fitness cost regime to fix in the simulations. We observe that in these states, there are no mutants with negative fitness effects in resistance level. Therefore, there is a dependency on time constants and fitness effects for the emergence of resistant mutants in our phase-space of TEM-1 resistance to cfx.

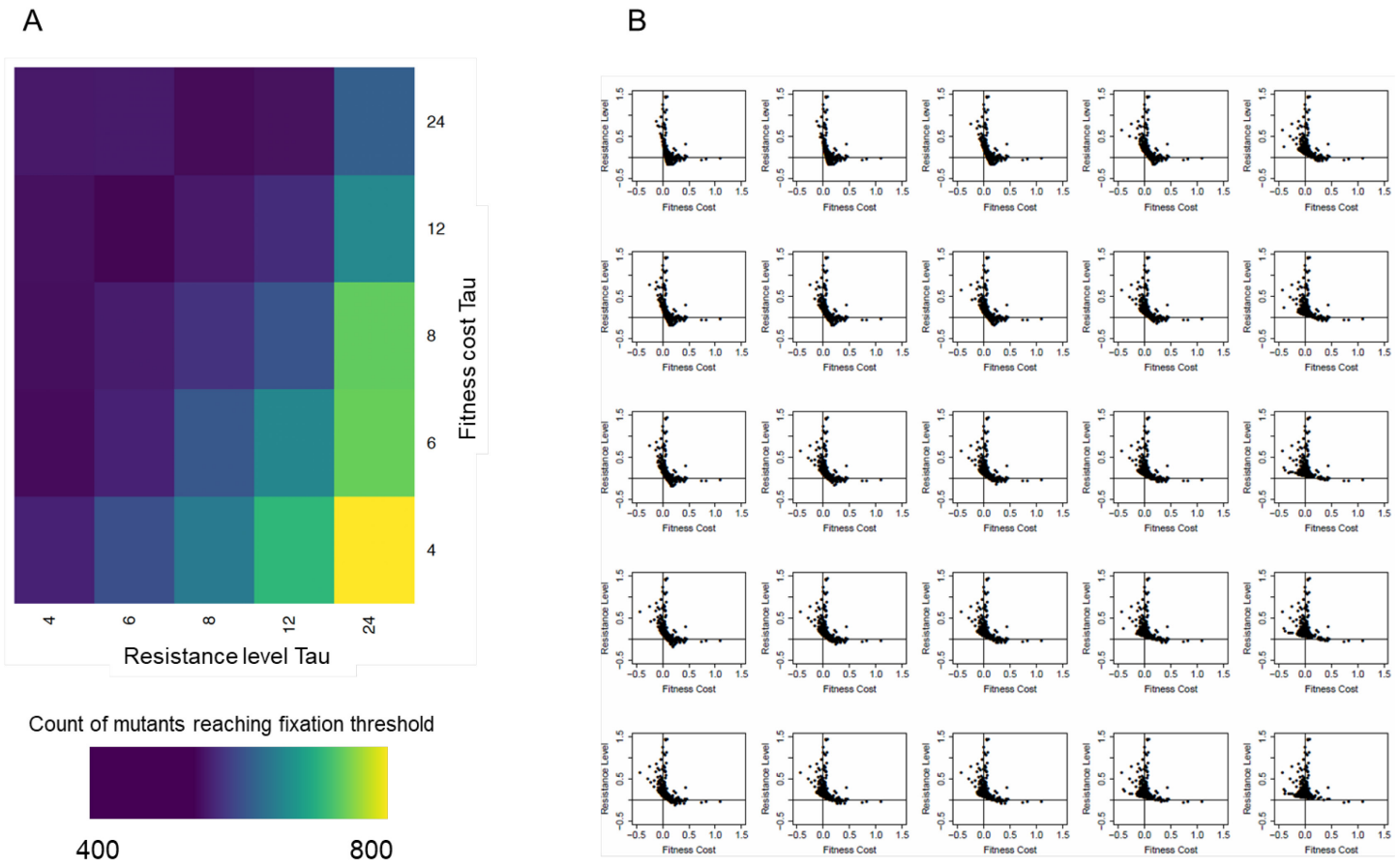

Figure 21. Phase-space of TEM-1 resistance to cefotaxime.
To construct the phase-space, simulations are performed by combining all the different time constants used. **A.** X-axis: Resistance level time constants. Y-axis: Fitness cost time constants. The phase-space of resistance of TEM-1 reveals a possible optimal treatment against the emergence of resistance by diminishing the overall count mutants in a population. **B.** The mappings demonstrate the states in the phase-space with the lowest number of mutants are characterized by high fitness effects in both the resistance level and fitness cost regimes. The composition of each state is shifted, depending on the combination of tau used. Each type of mutant has a dependency on a specific combination of tau.

Discussion

In this chapter, we further develop our initial evolutionary model, presented in Chapter 3. We seek to analyze evolutionary dynamics in finite populations by stochastic formulations. As the deterministic model of additive fitness cannot fully recapitulate the pathways observed in clinical isolates, we employ the stochastic Moran process to model our evolutionary dynamics (Nowak, 2006). We combine principally elements of population genetics in our model and specifically focus on identifying resistance-conferring single point mutations. In this context, instead of having all the mutants competing against each other, we only allow competition between a mutant strain and a WT strain. By doing so, we can identify the first passage time for the emergence of resistance by determining which mutations are viable to outcompete the WT strains in specific selective environments. Competition has been shown to be a potential key driver for selection of beneficial and deleterious mutations, as demonstrated by a previous dynamic evolutionary model (Osmond and de Mazancourt, 2013). We develop an evolutionary model that captures the probability of fixation of each mutant under different selection regimes. To summarize, we construct a comprehensive fitness landscape of TEM-1 under cfx selection from published DMS. We develop an evolutionary model based on the stochastic Moran process to determine the probability of survival for each mutant. We use the model to study the effects and fitness cost on resistance-conferring mutations.

We investigate the role of fitness cost in resistance and determine if it is possible to design optimal drug regimens as one of the strategies to alleviate resistance (Lee et al., 2013). In this context, we perform simulations with different time constants (tau) at constant drug concentration. Tau is an important parameter during antibiotic treatment as the total amount of time in a selective-environment determines the emergence of resistance (Paterson et al., 2016). To summarize, we use various oscillating treatments to determine an optimal drug dosage regimen. We construct the phase-space of resistance for TEM-1 under cfx resistance to further explore optimal dosage regimen. By using the stochastic model of evolution, we identify possible conditions to alleviate the effects of resistance.

Fitness effects dictate the mutants' evolutionary trajectories

The objective of the stochastic model is to recapitulate the growth behavior of each TEM-1 mutant compared to WT. As there is no competition in our simulations, fixation of a mutant is dependent on the fitness effects alone (Ueda et al., 2017). The model favors mutants with positive resistance level or fitness cost, but it will not always dismiss mutants with negative fitness effects in either regime. In the previous model, a mutant with a strong selection coefficient could be supplanted by another mutant with a slightly more beneficial selection coefficient (Nowak, 2006).

We constructed a total of 4997 evolutionary pathways under cfx selection. Each pathway represents the mutant's growth in a population over time. The pathways form three distinct clusters: mutants with growth only in the resistance regime; mutants with growth only in the selection-free regime; and mutants without growth in either regime. The variance in behavior is based on the different fitness effects in the resistance level regime and the fitness cost regime (Andersson and Hughest, 2010). The mutants in the first cluster are characterized by high fitness effects in the resistance level and are resistant to cfx treatment. The second cluster groups all mutants with positive fitness effects on the fitness cost regime only. These mutants are susceptible to cfx selection and cannot confer resistance in the presence of cfx. Our results parallel previous experimental findings which demonstrated that mutants with positive fitness effects in the fitness cost regime are favored to reach fixation in a selection-free environment (Knöppel et al., 2017). As demonstrated in a previous computational study (Torella et al., 2010), although these mutations do not confer resistance – in this case, cfx resistance, they could potentially lead to resistance to other antibiotics. The first two clusters represent key mutants which impact greatly their population and should be analyzed further in subsequent studies.

Identification of resistant mutants from literature

Next, we seek to determine if the fixed mutations returned from our evolutionary model are identified in the literature as clinical isolates. We define a mutant as resistant if its fraction

reached the designated fixation threshold. We stipulate that a mutant has a higher probability of surviving when it composes the majority of its population. Thus, the higher the fraction, the greater its probability to survive and possibly confer resistance. We compared all the mutants from our simulations against clinical isolates, as identified by Salverda et al. (2010).

If we do not consider resistance mutations that were observed in the literature only once, our model identifies 30% of the clinical isolates reported in **Table 3**. Thus, our stochastic evolutionary model can recapitulate a portion of known resistant mutants. Similar to the first model, it still cannot accurately recapitulate all the clinical isolates from the literature. This might be due to the simulations being performed for each mutant individually. Consequently, the model cannot recapitulate the complete gene interaction network of clinical isolates where a specific sequence or combination of mutations confers resistance (Shao et al., 2008). Although we can't use the model to study complete evolutionary pathways, this iteration of the model is better at studying the first passage time of resistance. Therefore, we should be able to identify the all-important first mutations of a pathway to resistance (Rodriguez-Verdugo et al., 2015). Consequently, we postulate that the model has a higher accuracy than the deterministic model to study resistance-conferring single-point mutations.

Fitness cost decreases the overall quality of resistant strains in our model's population

We then utilize the stochastic evolutionary model to study methods to alleviate resistance. We seek possibilities to reduce the number of mutants that could give rise to resistance. Initially, we concluded that fitness cost does not play an important role as its effects were negligible on the top common mutations from clinical isolates. This was accentuated by the fact that the effect of fitness cost can decrease for certain clinical isolates, rendering the effects neutral, as shown for fluoroquinolone resistance in *Pseudomonas aeruginosa* (Agnello et al, 2016). We believed that only the fitness effects of resistance level were of interest in the emergence of cfx resistance in TEM-1 since mutation M182T did not fix in our simulations due to neutral fitness effects in the resistance level. Thus, we seek to determine the relevance of fitness cost in our model.

We first test the model by removing the fitness cost regime from our simulations. By removing fitness cost, we seek to determine if there would be a change in the counts of mutants or the quality of the resistance-conferring mutants. When fitness cost is not included, we observe a sharp decrease in mutants that overtake their population. At first, we deemed this to be beneficial against resistance. If we remove fitness cost, there are fewer mutations and opportunities for resistance to emerge in a population. Thus, we initially stipulated that fitness cost effects do not purge the resistant mutants out of a population but could increase the overall number of mutants

We then seek to determine the quality of the mutants with increased survivability – ability to remain alive. Indeed, fewer mutants arising in a population would be better if the mutants do not impart resistance. Interestingly, for the simulations with fitness cost, the average resistance level is lower compared to the average resistance level of the simulations without fitness cost. By removing fitness cost from the simulations, fewer mutants are fixed in the population, but the mutants are fitter in the antibiotic environment. Consequently, a higher chance of selecting a mutant that confers strong resistance to the antibiotic would also increase overall fitness (Roux et al., 2015). By considering fitness costs, newly emerging mutant could outcompete a highly resistant mutant in a population, potentially reducing the number of resistance-conferring mutants in the gene pool. Competitive fitness assays have been used to show that fitter mutants would outcompete lesser fit mutants in selection-free environments (Melnyk et al., 2015). We stipulate that fitness cost decreases the overall quality of available resistance-conferring mutations, therefore there is an increased likelihood to fix a mutant with weaker fitness and decreased resistance. These findings agree with a theory previously proposed for the reversibility of resistance, where compensatory mutations should eventually restore WT fitness in the absence of antibiotics (Schulz zur Wiesch et al., 2010). Therefore, fitness cost is an important parameter to consider when designing an optimal drug dosage regimen.

Time spent in the regime with only fitness cost is key to alleviating resistance

To design an optimal drug dosage regimen, we construct the phase-space of TEM-1 resistance against cfx to study the effects of different tau and to determine the role of time of administration in the emergence of resistance (Lee et al., 2013; Paterson et al., 2016). The phase-space gives us a better understanding of the compensation system for resistance level and fitness cost while in-treatment. From each state of the phase-space, we observed that less time spent in the resistance level regime lowers the odds of gaining resistant mutants. Similarly, allowing more time in the fitness cost regime would lead to fewer resistant mutants arising. Therefore, we concluded that the best treatment against the emergence of cfx resistance in TEM-1 is a combination of a small tau in the resistance level regime and a high tau in the fitness cost regime. We can exploit this time-dependent behavior to potentially identify an optimal treatment to reduce the emergence of resistance in a population. Exploiting time of administration has been shown to diminish the emergence of resistance, although this was only tested on a localized setting and the fluctuations were in months instead of days or hours (Takesue et al., 2010). However, it is important to note that this kind of treatment could potentially lead to the emergence of other antibiotic-susceptible strains as we are still allowing the evolution of other mutants (Salvatore et al., 2016). Therefore, further studies in identifying an evolutionary balance against resistance will be necessary.

Summary of the stochastic evolutionary model

To summarize, we implement the stochastic Moran process in our evolutionary model to study the emergence of resistance. Even with the Moran model, we still cannot capture the full range of mutants from clinical isolates and laboratory evolution. This is expected as our model replicates the growth behavior for individual mutants only and it takes into consideration epistatic constraints. Although the stochastic model has proven to be moderately effective in the identification exercise, its accuracy still depends on the quality of the input dataset. Accuracy and balance of a dataset are key factors for optimal predictions (Vihinen, 2012). Nevertheless, we conclude that the stochastic model performs better than the deterministic one

to identify the first passage time of resistance. Therefore, we use the model to design optimal drug dosage regimens to alleviate the emergence of resistance. First, we determine that fitness cost might have a role in alleviating resistance by potentially allowing antibiotic-susceptible strains to fix in a population, thus diminishing the number of resistance-conferring mutations in the gene pool. Finally, we exploit fitness cost effects to design an optimal drug dosage regimen that would consist of a long fitness cost regime combined with a small resistance level regime. Thus, from our initial results, we determine that the stochastic evolutionary model could be used as a framework to study the emergence of resistance.

Chapter 5: Fitness landscape of TEM-1 for cefotaxime

In the previous chapters, we developed evolutionary models to identify the emergence of resistance. The stochastic model's accuracy still depends on the quality of the input dataset. Therefore, we seek to construct our fitness landscapes of TEM-1 to test our evolutionary models and to study the emergence of resistance. In this chapter, we determine the fitness landscape of TEM-1 under cfx selection. We first construct large-scale mutagenesis datasets of TEM-1. The comprehensive mutagenesis libraries are then used in competitive selection assays. As there is currently no dataset with the desired coverage for TEM-1 resistance to cfx, we perform the selection assays for a wide range of cfx concentrations. Also, we perform a continuous selection experiment where the library was first submitted to a selective environment and then switched to a selection-free environment, similar to the theoretical model developed by Ashcroft et al. (2014). Finally, following the selection assays, we use deep sequencing to determine the variants present in our libraries. Our experimental landscapes are then used in our stochastic evolutionary model to determine if we can analyze mutants' evolutionary behaviors with our own experimental dataset. Using our experimental approach with one of the TEM-1 subgroup, we found that our fitness landscapes were still incomplete. Thus, we need to optimize our approach before we can use our experimental fitness landscapes in our evolutionary models.

Methodology

TEM-1 WT recovery

The TEM-1 β-lactamase plasmid was obtained from the Whitehead group on Addgene (pSALECTNK-TEM1 (S70A, D179G), Plasmid catalog #81163). The plasmid was sent in an agar slab of DH5α bacteria. To rescue the plasmid, the bacteria were grown overnight at 37°C in 3 mL of LB with chloramphenicol (working concentration at 34 mg/mL). The plasmid encodes the TEM-1 sequence from H26 to W290. Two mutations are incorporated into the sequence: S70A and D179G. As we require the WT strain of TEM-1 to construct our fitness

landscapes, we recovered the WT strain through site-directed mutagenesis (QuikChange II Site-Directed Mutagenesis Kit, catalog # 200523). We designed two oligonucleotide primers with appropriate GC content percentage (40% to 60%) and melting temperature (T_m below 72°C). The primers used to revert the sites to WT are A70S (AGC), and G179D (GAT). The Agilent QuikChange protocol was used (Appendix 1). Sequences were confirmed by Sanger sequencing at IRIC.

TEM-1 library primers design

We used an in-house script to prepare the library primers for our comprehensive mutagenesis. The primers are used to substitute all 20 amino acids at every position of TEM-1. The script replicates saturation mutagenesis. The technique consists of randomizing a set of codons to produce all possible amino acids. The script requires the nucleotide sequence of TEM-1 and generates the library primers, replacing the targeted amino acid with an NNK degenerate codon. The NNK degenerate codon is used to create all our mutagenesis primers. All library primers have a G-C% content between 40% and 60%, a T_m between 53°C and 65°C, and a length between 27 and 40 bp. If possible, the NNK degenerate codon is located in the middle of the primer. To ensure a complete coverage of the TEM-1 coding region with the Illumina NextSeq 500 sequencing, we divided the TEM-1 coding sequence into four tiles (amino acid positions 25-91 (Tile 1), 92-156 (Tile 2), 157-222 (Tile 3), and 223-290 (Tile 4)). Each tile is subjected to antibiotic selection and sample preparation for Illumina sequencing. The library primers are ordered from IDT (Integrated DNA Technologies) in duplicate 96 Deepwell plates.

TEM-1 comprehensive whole-gene saturation mutagenesis library

Nicking Mutagenesis: comprehensive single-site saturation mutagenesis

Deep mutational scan (DMS) is a new method to assess the effects of thousands of mutant variations on a protein function through massively parallel functional screens and counting via deep sequencing (Fowler, 2014). To generate a DMS of a protein and construct its fitness

landscape, a high-quality mutational library of the query protein is required (Wrenbeck et al., 2016). There are several approaches to generate a comprehensive mutagenesis library. Currently, the most commonly used method is PFunkel. PFunkel offers the best library coverage, mutational efficiency, and control over the number of mutations introduced (Firnberg, 2012). However, PFunkel is limited by the requirement of an uracil-containing ssDNA template (dU-ssDNA). The main issue with dU-ssDNA is the highly variable yield of mutagenesis libraries. Therefore, we utilize the method introduced by Wrenbeck et al. (2016): plasmid-based one-pot saturation mutagenesis. This methodology is based on the versatility of nicking mutagenesis to generate comprehensive single-site saturation mutagenesis libraries. Nicking mutagenesis does not rely on dU-ssDNA and therefore has a higher yield reliability compared to PFunkel (Wrenbeck, 2016). Another key advantage of the nicking mutagenesis methodology is the possibility to prepare a comprehensive mutagenesis library in a single day.

Nicking mutagenesis is a flexible methodology that can be performed on any plasmid double-stranded DNA (dsDNA) if it contains a 7-bp BbvCl restriction site. Nicking mutagenesis can generate the comprehensive single-site saturation mutagenesis library for any protein as long as it contains the restriction site. The restriction site is recognized by a pair of endonucleases, Nt.BbvCl and Nb.BbvCl, that each nick the same site on the dsDNA. This nicking is primordial to successively create and degrade a wild-type ssDNA template to introduce the mutagenesis primers. The methodology is divided into four major steps, as shown in **Figure 22**. The generated library is then used in selection assays to construct the fitness landscapes of the protein of interest.

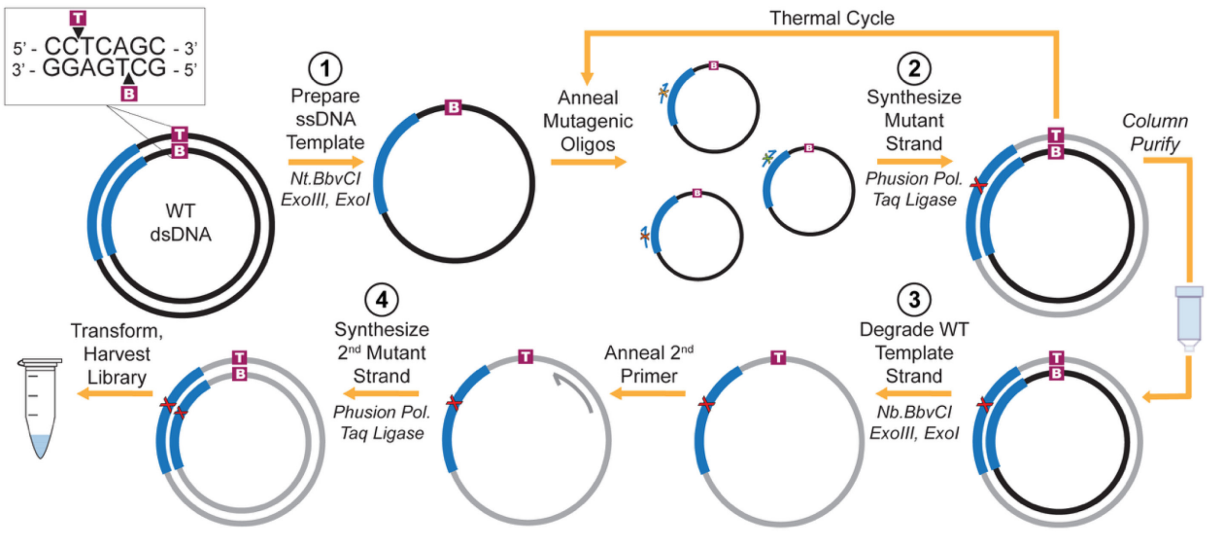

Figure 22. Nicking mutagenesis to construct a mutagenesis library.
1. The dsDNA plasmid is nicked via a strand-specific site by Nt.BbvCl to create a ssDNA template. The nicked strand is then selectively digested by exonuclease III. **2.** To synthesize the mutant strand, the DNA template is thermal-cycled with mutagenic oligos at a low primer-to-template ratio to ensure appropriate annealing of each primer to each template. The primer is extended around the circular DNA template with high-fidelity Phusion DNA polymerase. Taq DNA ligase finally closes the new strand to form a heteroduplex dsDNA with a mismatch at the targeted mutational site. **3.** Nb.BvCl, the opposite-strand nicking endonuclease, creates a nick in the template strand which is degraded by exonuclease III. A ssDNA with the mutated site is created. **4.** A secondary primer is added and synthesizes the complementary mutant strand to resolve the heteroduplex and have a dsDNA which encodes a single point mutation. (Adapted from Wrenbeck et al., 2016)

The main steps of our protocol to generate our TEM-1 mutagenesis libraries are presented here:

1. 0.76pmol of the recovered TEM-1 WT plasmid is used as the template for the library.
2. The plasmid contains the 7-bp BbvCI recognition site that is first nicked by Nt. BbvCI to generate the ssDNA template. The nicked strand is selectively digested by ExoIII (exonuclease III).
3. The designed mutagenic primers are then added at a 1:100 ratio with the template to synthesize the mutant strands by thermal cycling. The primer is extended around the circular ssDNA template with HF Phusion DNA polymerase (high-fidelity Phusion DNA polymerase).

4. Taq DNA ligase is added to the PCR cycle to close the new mutant strand to form a dsDNA plasmid with a mismatch at the mutational site.
5. The WT template strand is nicked then by Nb.BbvCI, the opposite-strand nicking endonuclease, and degraded by ExoIII to create a ssDNA of the mutant strand.
6. A secondary primer is added to synthesize a complementary mutant strand to yield mutagenized dsDNA.
7. The plasmids are transformed into XL-1 Blue competent cells (Agilent) in square BioAssay dishes containing LB with tetracycline and chloramphenicol. The libraries are then incubated overnight at 37°C overnight.
8. The libraries are scraped from the square dishes and conserved into 1.5mL glycerol stocks in -80°C.

The exact experimental protocol is presented in Appendix 2.

TEM-1 library selection assays

To determine the appropriate cfx concentrations to use in the selection assays, we performed growth rates experiments of TEM-1 WT. The WT strain was grown overnight in selective media (LB + Chl + Tet) until saturation at 37°C. A MicroWell 96-Well Microplate was prepared with the different concentrations to assess. Wells contained concentrations of either 0,06225, 0,125, 0,25, 0.5, 1, 2, 4, 8, 16, and 32 µg/mL cfx. The overnight culture is diluted into each well to obtain a starting OD_{600} of 0,1. The cultures were then grown at 37°C with constant shaking at 225rpm for 24 h. Growth for all the wells was monitored at every hour in the Tecan Spark. Following our initial growth rate assays on TEM-1 WT strains, we determined that our libraries should be grown in sub-MIC cfx concentrations.

For cfx selection assays of the TEM-1 mutagenesis libraries, we diluted each library in pre-warmed LB until OD_{600} = 0.1. Each culture was transferred to growth tubes that contained LB at concentrations of 0, 0,02, 0,2, 1, or 4 µg/mL cfx. The colonies were grown at 37°C with constant shaking at 225rpm for approximately 4 h. OD_{600} was monitored for all tubes at every hour. The tubes are then transferred to ice to immediately stop growth, washed of antibiotics

by centrifugation and resuspended in non-selective LB. The plasmids were purified by miniprep and prepared for Illumina sequencing.

TEM-1 Illumina sequencing

For each concentration, all plasmid samples at t = 0 h and t = 4 h were prepared for Illumina sequencing by PCR, as performed by Kowalsky et al., (2015). The adaptor sequences for Illumina sequencing were added in one single PCR round. The adaptor sequences add the annealing site for the Illumina paired-end sequencing primers and a 6 bp barcode for multiplex sequencing. The barcodes are generated from the RPI-series primers obtained from Kowalsky et al. (2015). They are partially derived from the Illumina RNA TruSeq preparation kit. The samples were purified by gel purification before being sequenced to confirm adequate annealing of all primers. NGS was performed at IRIC's Genomics Platform on an Illumina NextSeq 500. Sequences were cycled by paired-end read with a maximum pairing of 2 x 155nt for each with a sequencing depth of 4 million reads per sample. The protocol to prepare the sequences for NGS is presented in Appendix 3.

NGS analysis

The NGS analysis consists of a set of custom Python scripts. The scripts are used to extract the single mutants and wild-type sequences from the R1 and R2 fastq files. First, the R1 and R2 files are merged with USEARCH's fastq_mergepairs command. The consensus nucleotide sequence between R1 and R2 with the highest quality score is selected. Reads with sequencing errors or no nucleotides detected (N's) are dismissed from the initial parsing. The nucleotide sequence is then translated to its amino acid sequence. WT sequences and sequences with single amino acid changes are kept to construct the fitness landscapes of TEM-1 under cfx selection. The relative fitness effects for each mutant is calculated with **Eq. 1**. The dataset is then formatted appropriately in a CSV file to use in our stochastic evolutionary model.

$$F_i^a = \log_{10}\left(\frac{N_i^{a,sel}}{N_i^{a,unsel}}\right) - \log_{10}\left(\frac{N_i^{wt,sel}}{N_i^{wt,unsel}}\right) \quad \text{(Eq. 1)}$$

Where the relative fitness F_i^a of each amino acid mutation *a* at each position *i* is determined as the logarithm in the allele counts (*N*) between the selected population ($N_i^{a,sel}$) and the unselected population ($N_i^{a,unsel}$), relative to the wild-type allele.

Results

To determine the efficiency of our experimental approach, we construct a partial TEM-1 fitness landscape with the first subgroup of our mutagenesis library (Tile 1). This allows us to test out our experimental approach with one of the TEM-1 subgroups before performing it on all 4 subgroups.

TEM-1 selection assays

We analyze the growth assays of WT to determine the appropriate cfx concentrations to use in our selection assays. We observe growth of WT strains at cfx concentrations lower than 4 μg/mL, which corresponds to sub-MIC concentrations (Olofsson et al., 2005). Growth is unaffected in samples containing 0,5 μg/mL or less. At cfx concentrations higher than 4 μg/mL, no significant growth was detected for WT. Therefore, we perform our selection assays at cfx concentrations of 4, 1, 0,2, 0,02 and 0 μg/mL. The growth assays are presented in **Figure 23**. We observe regular growth for our libraries containing 0,02 μg/mL cfx or less. Growth was hampered on the samples containing 0,2 μg/mL cfx or more. For the NGS analysis, we use samples from Tile 1 at 0 μg/mL (C0) and at 0,2 μg/mL (C02), extracted after 3 h 30. An additional sample was taken at time 0 h (T0). This sample is used in our calculations of relative fitness effects.

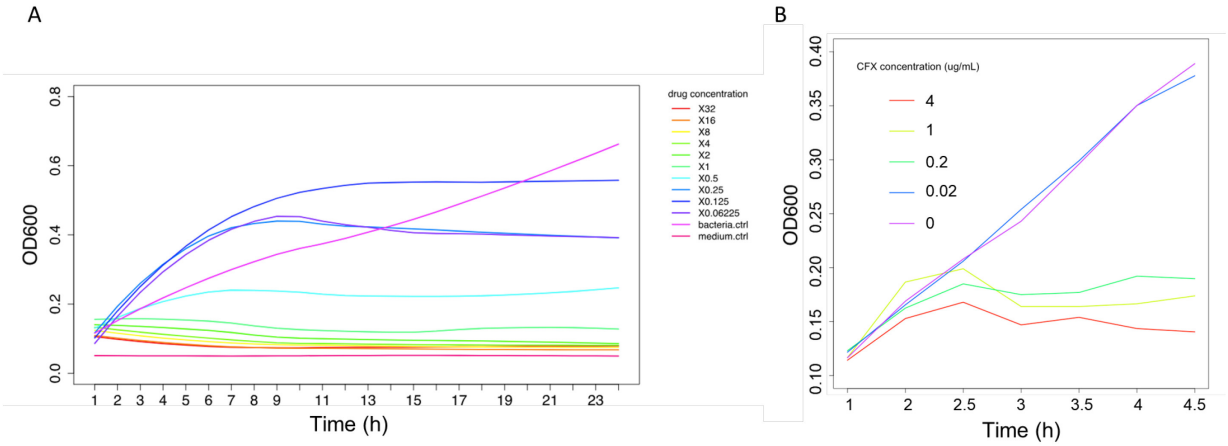

Figure 23. TEM-1 selection assays for WT and libraries
We perform an initial growth experiment to determine cfx concentrations which did not hamper growth. The concentrations identified also had to be sufficiently high to exert selective pressure. **A.** We perform the growth assays for a total of 24 h with a WT strain of TEM-1, ranging from 32 µg/mL to 0 µg/mL cfx. We observe that for concentrations higher than 4 µg/mL cfx, samples had difficulty growing (n=3). **B.** Selection assays on Tile 1 of our TEM-1 libraries were performed for 4,5 h. Samples were taken after 3,5 h.

NGS analysis

Following NGS, we obtain a total of 1 552 603 reads for Tile 1 at T0. For the selection assay samples, we obtain a total of 3 618 985 reads and 1 535 999 reads for C0 and C02, respectively. After filtering all trimmed paired-end sequences and merging the sequences with USEARCH's fastq_mergepairs command, we obtain a final number of 937,766 reads, 2 111 323 reads, and 864 689 reads for T0, C0, and C02 respectively. From the extracted sequences, we determine the number of sequences that are only one mutation away from the WT. As shown in **Figure 24A**, a significant amount of sequences, around 50% of them, have more than one mutation and are removed from the analysis. From **Figure 24B**, we observe we do not have a complete coverage of Tile 1 as there is a considerable amount of missing mutations, most notably at sites V31, P62, S70, and L91. Thus, our whole-gene saturation mutagenesis library for Tile 1 does not cover all possible mutations. We generate the partial distribution of relative fitness effects of TEM-1 with and without selection in **Figure 24C**. We

observe that both distributions are primarily composed of mutations with positive fitness effects, and a large fraction of them have neutral or near-neutral effects.

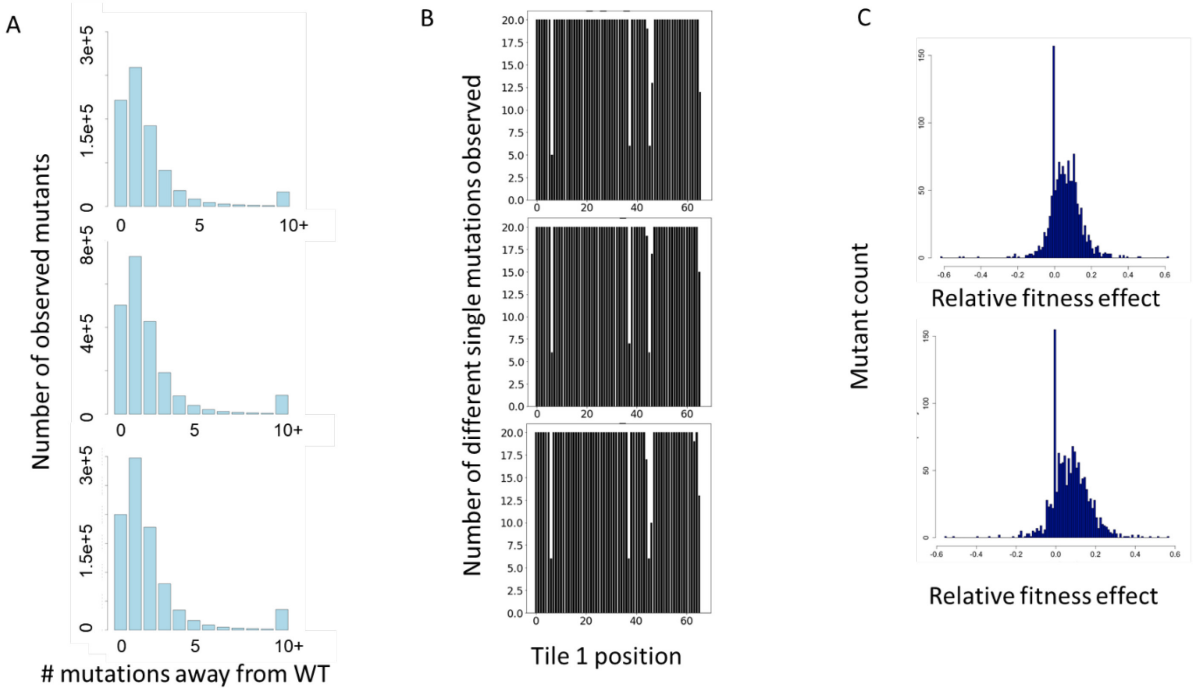

Figure 24. NGS analysis of TEM-1 under cfx selection.
Post-analysis from the sequencing results returned from IRIC. Samples for analysis are the T0, C0, and C02, respectively, from top to bottom. **A.** We determined the number of mutants with a specific number of mutations away from the WT. Only WT and single-site mutants are of interest in the construction of the TEM-1 fitness landscape. **B.** We determined the total number of different single mutations observed to ensure complete coverage of our library. This is not the case due to several missing key mutations. **C.** We calculate the relative fitness effects of two datasets, one without selection (top) and one with (bottom), as defined by Eq. 1 where samples at t = 0 h are assigned as the $N_i^{a,unsel}$ and samples at t = 5 h are assigned as the $N_i^{a,sel}$.

To ensure that the two TEM-1 fitness landscapes are different. We calculate the Pearson correlation between our fitness landscapes, as shown in **Figure 25**. We observe a strong correlation (0,739) between the two experimental fitness landscapes. Finally, we observe no

correlation (0,06) between our fitness landscape of TEM-1 under cfx selection and the reference fitness landscape of TEM-1 under similar selective conditions (Stiffler et al., 2015).

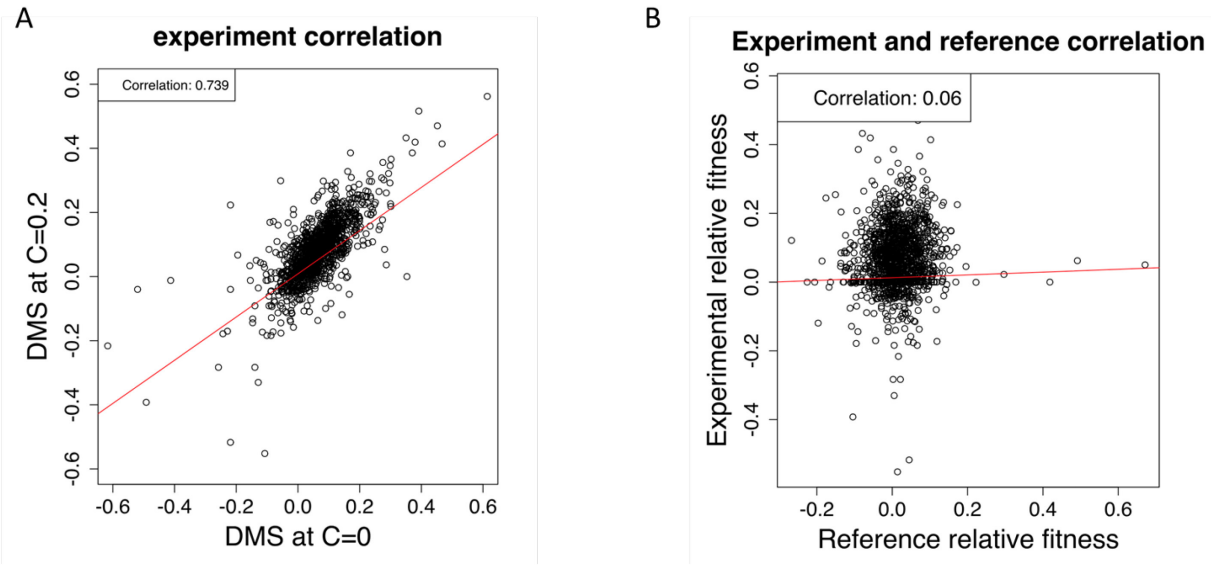

Figure 25. Correlation between experimental fitness landscapes.
We use the Pearson correlation to determine if there is a significant relationship between our experimental DMS. We calculate the log enrichment for each mutant. **A.** We determine the correlation between the DMS with selection and without selection. There is a strong correlation between the two datasets as the dynamic range is not sufficient to differentiate the allele frequency to calculate the log enrichment. The correlation we observe between the two datasets is from the T0 dataset used to calculate the relative fitness effects when using Equation 1. **B.** We observe no correlation between our datasets and the reference dataset from Stiffler et al. (2015).

Simulations with our own dataset

Using our experimental fitness landscapes, we want to determine if we could accurately simulate the evolutionary behavior of TEM-1 during a continuous selection assay. We perform a selection experiment in which our Tile 1 mutagenesis library is initially in a selective environment and then switched to a selection-free environment. Tile 1 was submitted to 4 h of selection regime at 0,2 µg/mL cfx and switched to an environment without selection for the next 4 h. Samples were taken at 0 h, 4 h, and 8 h. We construct the experimental fitness

landscapes specific to the continuous experiments, where the change in environment is immediate and there is no interruption in the selection assays. The fitness landscapes are presented in **Figure 26**. The single selection assays are used in our simulations.

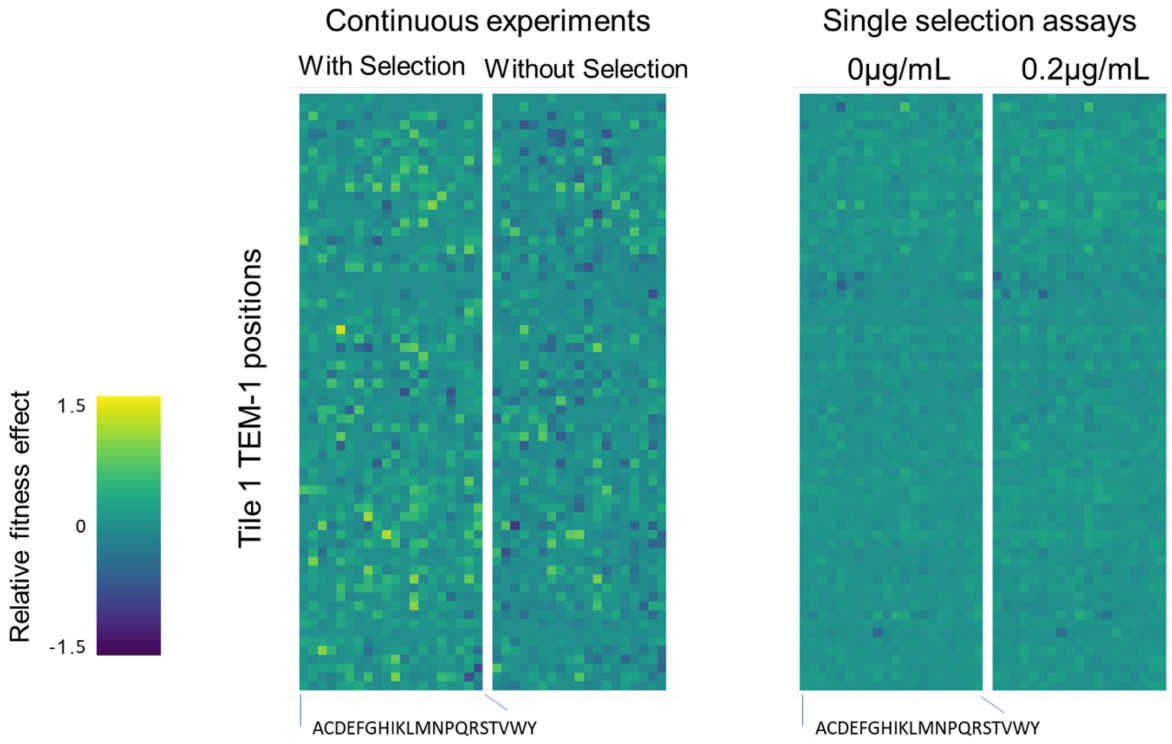

Figure 26. DMS datasets of TEM-1 from our continuous experiments and single selection assays.
DMS obtained from our selection assays. The continuous DMS were obtained from the on/off experiments while the single selection assays were obtained from the libraries subjected to a single selective environment. Each row represents a TEM-1 position, and each column is one of the twenty possible amino acid mutations. Beneficial mutations are highlighted in yellow while deleterious mutations are highlighted in blue. The DMS from the single selection assays will be used in our simulation to assess whether it is possible to replicate the results obtained from the continuous experiments.

We perform the simulations with our stochastic evolutionary model. The same conditions are used than in previous simulations, the only difference is the tau and total time *t* used. To replicate the experimental conditions from the continuous assay, we set the tau to 4 h, and the *t*

to 8 h. Thus, the simulations are performed for 8 h only. We use the C0 dataset as the fitness cost regime and the C02 dataset as the resistance level regime. The fraction of mutants is recorded after 8 h. We generate a total of 1260 evolutionary pathways for all TEM-1 mutants in Tile 1. Finally, we compare the derived probability of survival obtained from the stochastic simulations against the relative fitness effects from our continuous assays.

We observe no correlation between the two experimental fitness effects and the derived probability of survival (-0,01), as shown in **Figure 27**. A mutant with a high probability of survival in our simulations does not have a large relative fitness effect in our continuous experiments. We compare the two datasets by rank-ordering the most beneficial mutations and most deleterious mutations. We observe no match between the mutants from the simulations and the experiments, as shown in **Table 6**.

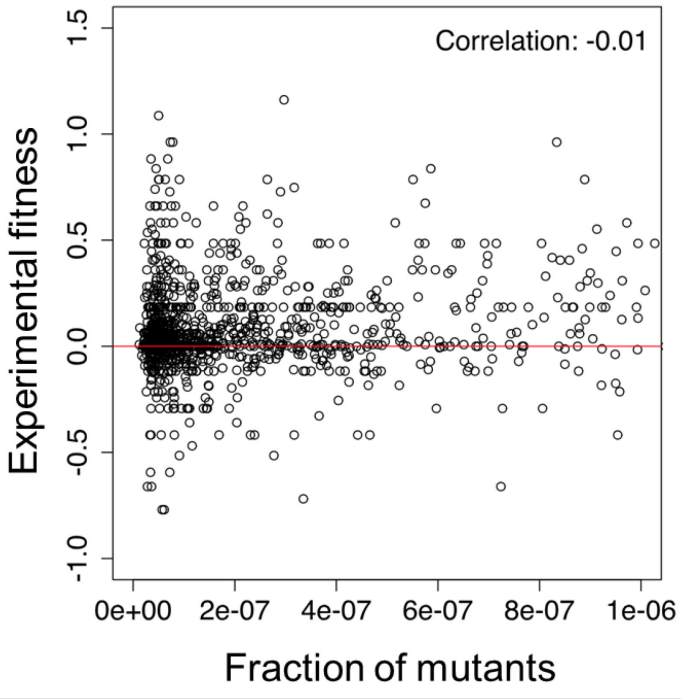

Figure 27. Correlation between relative fitness effects and probability of survival. We use the Pearson correlation to determine if there is a significant relationship between relative fitness effects from our experimental DMS, and the fraction of mutants from our simulations. From our simulations, we observe no correlation between the probability of survival and the relative fitness effects.

Simulation		Experiments	
Best mutation	Fraction of population	Best mutation	Fitness effect
I95A	0.047254584	M68H	1.297892
R83E	0.001243048	V44I	1.1616729
R83V	0.000512608	Y46I	1.087039
R83C	0.000403652	L30H	0.9621996
Worst mutation	**Fraction of population**	**Worst mutation**	**Fitness effect**
L40N	1.2e-08	E28A	-0.7702931
L40K	1.3e-08	T29D	-0.7702931
P62C	1.8e-08	R61H	-0.7191406
G62Q	1.8e-08	R61N	-0.6611486

Table VI. Mutation comparison between simulation and experiments.
We compare the mutations count from the end of the simulations against the relative fitness from experiments. From the simulations, we observe that a mutant that has obtained a high probability of survivability is not reflected in its fitness effects obtained from the experiments. There is no correlation between the two.

Discussion

In this chapter, we perform DMS of TEM-1 at different cfx concentrations. Our DMS generate large-scale mutagenesis datasets of TEM-1 for each of its mutation variants under selective conditions. We use a methodology that is based on the nicking mutagenesis protocol proposed by Wrenbeck et al. in 2016, and the selection assay protocol proposed by Stiffler et al., in 2015. We construct the fitness landscapes to profile the functional effects of TEM-1 under cfx selection. As we want to compare our fitness landscapes with Stiffler et al. (2015) TEM-1 fitness landscapes, we focus our initial analysis on two of our fitness landscapes: TEM-1 without selection, and TEM-1 with 0,2 μg/mL cfx selection. We also include experimental datasets sampled from a continuous selection assay. Our experimental fitness landscapes are then used in our stochastic evolutionary model to determine if we can predict mutants' evolutionary behaviors under selective pressure. We observe that our initial TEM-1 fitness landscapes are still incomplete and thus cannot be used in our evolutionary models.

Our TEM-1 fitness landscape is incomplete

Following NGS, our DMS contained a large number of mutants with more than one amino acid mutation in their sequence. We also have a considerable number of missing mutations. The source of multi-site mutants could originate from our experimental approach. The design of our saturation mutagenesis primers might not have been optimized for the nicking mutagenesis method. This combination of multi-site mutants and missing mutations significantly diminishes the overall size of our dataset and therefore, the quality of our fitness landscapes.

Our fitness landscapes are also skewed towards mutations of positive fitness effects. A substantial portion of our mutations is observed to be beneficial. This is contrary to the fact that a majority of mutations are neutral or near neutral. Only a small fraction of mutations are deleterious, and an even smaller amount of mutations are beneficial (Loewe, 2010). The fitness landscape with overall positive fitness effects can be explained by the difference in allele frequency and the small dynamics range observed in our samples. Consequently, the

difference in allele frequency between the two timepoints to calculate the relative fitness effects is not significant enough to observe a diverse range of fitness effects. If the range was larger, we would be able to observe a wide distribution of fitness effects.

Our TEM-1 fitness landscape does not cover a large dynamic range

The small dynamic range of allele frequencies observed in our DMS is one of the main explanations behind the strong correlation observed between our experimental fitness landscapes of TEM-1. The small dynamic range makes it difficult to differentiate the allele frequencies in mutants after selection. Although there is a significant difference in OD_{600} measurements between the two samples, the difference in OD might be explained by the WT strains growing at a much faster rate than their mutant counterparts. The mutants could be growing at a slower rate, thus 4 h is not sufficient to allow for us to detect mutant growth and quantify it.

Therefore, it is expected that there would also be no correlation between our experimental fitness landscape and the reference fitness landscape. The main difference is that their libraries were allowed to grow overnight until saturation. This also means that their libraries experienced significant fitness cost effects. Although this would ensure a dynamic range large enough to observe the differences in allele frequencies, it biases the DMS to fitness cost effects. In our DMS, we stop the growth of our colonies immediately following the selection assays. They are not allowed to grow overnight. This ensures that our samples are free from fitness cost effects, but it also means we have a lower allele count frequency in our DMS. The lack of correlation could also be explained by the overall positive fitness effects observed in our mutations. Mutants with negative relative fitness only compose around 10% of our fitness landscape, whereas the reference fitness landscape is composed of 50% of mutants with negative relative fitness. Thus, there is an overrepresentation of mutants with positive fitness effects in our fitness landscape. The lack of dynamic range and the overrepresentation of positive fitness effects will bias our evolutionary simulations.

No correlation between our fitness landscapes following simulations

We performed simulations with our experimental fitness landscapes to determine if we could recapitulate the mutants observed in our continuous experiments. From our experiments and simulations, there does not seem to be a correlation between survival probability and fitness effects. Neither the most beneficial or most deleterious mutations from the continuous experiments are identified in the simulations. These conclusions are expected as our DMS does not cover the necessary dynamic range required to make accurate observations.

Issues to solve for the TEM-1 fitness landscape

These simulations demonstrate the reliance of the model on the quality and reliability of fitness landscapes. There are still a few key issues to solve before generating our own fitness landscape suitable to use in our evolutionary model:

- Ensure that our libraries fully cover the required sequence map.
- Ensure that our libraries also have a sufficiently large frequency of different alleles.
- Ensure that our DMS covers a wide dynamic range without having the samples submitted to fitness cost effects. This is the most complicated issue to solve now.

It is key to determine the dynamic range of the whole library that is being tested, and not just of WT itself. Solving these issues will move us closer to generate an accurate fitness landscape to study resistance and fitness cost.

Chapter 6: General discussion

In this memoir, we study the mechanism surrounding the evolution of antibiotic resistance. We attempt to identify the potential ways to alleviate resistance by first developing a deterministic evolutionary model of additive fitness. The model combines population dynamics and the biochemical effects of mutations to capture the effects of purifying selection under selection with Amp. We further develop the initial model by implementing the Moran process. The stochastic model of microbial evolution aims to determine mutations that might be enriched during antibiotic treatment. Here, we compare the difference between the stochastic and deterministic models.

Stochastic model vs deterministic model

We first construct a deterministic evolutionary model by having all TEM-1 mutants compete against each other in the same environment. Then, we further develop on the initial model by implementing the Moran process for stochastic evolution. With the stochastic model, we aim to determine the evolutionary trajectories of individual mutants. The Moran model does not incorporate competition between multiple mutants, instead, the mutants only compete against the WT strain. One of the key advantages of the Moran model is that a beneficial mutation is not always guaranteed to reach fixation. There is always a risk of extinction for the mutant. In a deterministic model, given enough time, an advantageous mutation, regardless of how small the advantage over its competitor, will always fix in a population (Nowak, 2006). Thus, the stochastic model provides valuable insight into the process of mutant selection in a population.

In the context of adaptive selection, contrary to the deterministic model, our stochastic model can accurately capture a large portion of the relevant mutants identified from clinical isolates and laboratory evolution experiments. The stochastic model potentially has a higher accuracy than the deterministic one if the data is accurate and reliable (Vihinen, 2012). This model can be used to further study the role of fitness cost in resistance while in treatment, and to design optimal dosage regimen to alleviate the effects of resistance.

Fitness cost and its role in alleviating resistance

Many resistance-conferring mutations in bacteria impose a fitness cost to the carrying organism to maintain the mutation (Melnyk et al, 2015). Based on this observation, it's suggested that one of the best treatments to alleviate resistance, is to stop or reduce the usage of antibiotics (Maharjan et al., 2017). This approach has not always proven to be efficient. Clinical studies have demonstrated that, in some cases, despite the absence of antibiotics, the resistant bacteria remained abundant in the population (Sundqvist et al., 2010) or even increased in frequency (Arason et al., 2002). The known examples for this are for *E. coli* against trimethoprim-containing drugs, and *Streptococcus pneumonia* against penicillin. Other groups reported the expected behavior of a decrease of resistant bacteria in the population (Gottesman, 2009). It was also observed that the reduced use of antibiotics rarely succeeds in completely eliminating resistant strains (Enne, 2010). Therefore, it remains unclear how fitness cost affects antibiotic resistance. We use the model to study the effects of fitness costs in resistance and to determine if it is possible to exploit fitness cost to alleviate resistance.

Using the stochastic model, we seek to design an optimal drug dosage regimen that could alleviate the emergence of resistance. we first determine the importance of fitness cost and its role in the emergence of resistance. From the simulations performed the DMS of TEM-1 under cfx selection, we identify one potential key role of fitness cost while in-treatment. Although fitness cost cannot eliminate the resistant strains from the population, it reintroduces potentially antibiotic-susceptible strains back in the pool of available mutants. As there is more competition against the resistant strains in the population, this diminishes their overall survival probability. Although the effect is not substantial, it is a crucial step to study potential methods to alleviate resistance.

Phase-space of resistance to design an optimal drug dosage regimen

To determine the possibility to alleviate drug resistance, we construct the phase-space of TEM-1 resistance against cfx. We combine different time constants for the resistance level regime and the fitness cost regime for the duration of an antibiotic therapy. Each state in the

phase-space is used to design an optimal dosage regimen to decrease the number of potentially resistant mutants. By exploiting the effects of fitness cost, we identify a potential oscillating regimen that alleviates the emergence of resistance. Thus, the model can be used for novel drugs to design a prescribed regimen to delay the onset of resistance in patients.

Limitations in our current evolutionary models

There are a few shortcomings to the stochastic evolutionary model:

- Currently, the model does not account for extinction in its calculation of survival probability. It will always assume there is either one surviving mutant or WT strain. Most of the evolutionary experiments calculate growth rate. This is satisfactory to simulate adaptive selection but to accurately model extinction, the data must also incorporate death rate. This is not the case for the current datasets we are using.
- The models do not emphasize important biophysical constraints such as epistasis (Serohijos et al. 2014) and protein stability (Dasmeh et al. 2014). For example, the model currently assumes that neither deleterious or beneficial mutations perturb protein stability. We have considered mitigating epistatic effects in our evolutionary models by only analyzing the first-step mutations to resistance. Therefore, these biophysical constraints are important to fully implement in subsequent versions of the evolutionary models.
- In the context of antibiotic resistance, the model does not take into consideration compensatory mutations. These mutations can arise after resistance to alleviate the fitness costs of resistance (Schulz et al., 2010). As such, the model only analyzes single-site mutations to determine the emergence of first-time resistance.
- The model should also incorporate clonal interference and complex dynamics of selection in microbial evolution (Good et al., 2017) as we want to eventually model long-term adaptation in constant selection environments.
- For optimal drug therapy design, the model doesn't take into consideration the pharmacodynamics and pharmacokinetics of the antibiotics in question. The efflux rate of an antibiotic is key to determine the fitness cost effects of a resistance mutation, and

the different rates can affect persistence and dissemination of antibiotic resistance (Olivares et al., 2014).

Improving the proposed model by considering these points would further increase accuracy and predictive power.

Chapter 7: Conclusions

Despite the shared theme of population demography between biophysics and population genetics, the two are rarely integrated together to study microbial evolution. A few theoretical models implementing both fields have been utilized to study evolution, but none have been used to study the emergence of resistance. Consequently, the application of evolutionary models and theories to the resistance problem is largely unknown. Therefore, the role and contribution of molecular biophysics and population genetics to the emergence of resistance remains unclear. In this memoir, we implemented evolutionary models considering principally population genetics constraints. We used concepts from theoretical evolutionary studies to investigate the emergence of resistance. The evolutionary models that were developed in this memoir will be the base of the multiscale model for the prediction of microbial evolution to study resistance in a single unified model. Further integration of complex biophysical constraints such as protein folding, will contribute significantly to increase the predictive accuracy of our evolutionary models. Understanding these complex biophysical constraints has been part of the Serohijos group's interest. Consequently, a significant amount of research in the group has been focused on elucidating epistasis using biophysical and population genetics approaches. This research resulted in two important findings on epistasis. It was determined that a significant amount of amino acid substitutions would have experienced epistasis due to simple selection for folding stability, accordingly linking epistasis to the strength of molecular selection (Dasmeh and Serohijos, 2018). Also, a proteome-wide scan in *E. coli* revealed that epistasis is stronger among highly expressed genes, therefore highlighting the combination of selection and epistasis in long-term evolution (Dasmeh et al., 2017). Combined with the recent advancements by the Serohijos group, the evolutionary perspective provided in this memoir will be an important foundation for research in antibiotic resistance.

Finally, to summarize, we present three key conclusions from this memoir:

1. We can identify the first passage time of resistance with our models, albeit with limited precision only. The accuracy of the prediction depends on the quality of the data.

2. We can design optimal drug regimens to mitigate the emergence of resistance, or to slow down resistance by exploiting fitness cost. This must be combined with diminished administration of antibiotics to obtain optimal results.
3. There is a possibility to study evolutionary pathways of TEM-1 to cfx resistance by using fitness landscapes generated from deep mutational scan data. Currently, we can only reveal limited information on the pathways due to the epistatic constraints of the model. The complete identification of evolutionary pathways to resistance remains a daunting challenge that will require additional development of the model and experimental methodologies.

Bibliography

Abel, Sören, Pia Abel zur Wiesch, Brigid M. Davis, and Matthew K. Waldor. 2015. "Analysis of Bottlenecks in Experimental Models of Infection." Edited by Deborah A. Hogan. *PLOS Pathogens* 11 (6): e1004823. https://doi.org/10.1371/journal.ppat.1004823.

Agnello, Melissa, Steven E. Finkel, and Annie Wong-Beringer. 2016. "Fitness Cost of Fluoroquinolone Resistance in Clinical Isolates of Pseudomonas Aeruginosa Differs by Type III Secretion Genotype." *Frontiers in Microbiology* 7 (October). https://doi.org/10.3389/fmicb.2016.01591.

Allen, Rosalind, and Bartłomiej Waclaw. 2016. "Antibiotic Resistance: A Physicist's View." *Physical Biology* 13 (4): 045001. https://doi.org/10.1088/1478-3975/13/4/045001.

Andersson, Dan I. 2003. "Persistence of Antibiotic Resistant Bacteria." *Current Opinion in Microbiology* 6 (5): 452–56.

Andersson, Dan I. 2006. "The Biological Cost of Mutational Antibiotic Resistance: Any Practical Conclusions?" *Current Opinion in Microbiology* 9 (5): 461–65. https://doi.org/10.1016/j.mib.2006.07.002.

Andersson, Dan I., and Diarmaid Hughes. 2010. "Antibiotic Resistance and Its Cost: Is It Possible to Reverse Resistance?" *Nature Reviews Microbiology* 8 (4): 260–71. https://doi.org/10.1038/nrmicro2319.

———. 2014. "Microbiological Effects of Sublethal Levels of Antibiotics." *Nature Reviews Microbiology* 12 (7): 465–78. https://doi.org/10.1038/nrmicro3270.

Andrews, Christine A. 2010. "Natural Selection, Genetic Drift, and Gene Flow Do Not Act in Isolation in Natural Populations." *Nature Education Knowledge* 3 (10): 5.

Arason, Vilhjalmur A., Adalsteinn Gunnlaugsson, Johann A. Sigurdsson, Helga Erlendsdottir, Sigurdur Gudmundsson, and Karl G. Kristinsson. 2002. "Clonal Spread of Resistant Pneumococci Despite Diminished Antimicrobial Use." *Microbial Drug Resistance* 8 (3): 187–92. https://doi.org/10.1089/107662902760326896.

Ashcroft, P., P. M. Altrock, and T. Galla. 2014. "Fixation in Finite Populations Evolving in Fluctuating Environments." Journal of The Royal Society Interface 11 (100): 20140663–20140663. https://doi.org/10.1098/rsif.2014.0663.

Bergen, Andrew C. 2015. “Mutation Load under Additive Fitness Effects.” *Genetics Research* 97 (February): e2. https://doi.org/10.1017/S0016672314000226.

Bergman, M., S. T. Nyberg, P. Huovinen, P. Paakkari, A. J. Hakanen, and and the Finnish Study Group for Antimicrobial Resistance. 2009. “Association between Antimicrobial Consumption and Resistance in *Escherichia Coli*.” *Antimicrobial Agents and Chemotherapy* 53 (3): 912–17. https://doi.org/10.1128/AAC.00856-08.

Bermingham, Alun, and Jeremy P. Derrick. 2002. “The Folic Acid Biosynthesis Pathway in Bacteria: Evaluation of Potential for Antibacterial Drug Discovery.” *BioEssays* 24 (7): 637–48. https://doi.org/10.1002/bies.10114.

Bershtein, Shimon, Korina Goldin, and Dan S. Tawfik. 2008. “Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins.” *Journal of Molecular Biology* 379 (5): 1029–44. https://doi.org/10.1016/j.jmb.2008.04.024.

Bershtein, Shimon, Adrian WR Serohijos, and Eugene I Shakhnovich. 2017. “Bridging the Physical Scales in Evolutionary Biology: From Protein Sequence Space to Fitness of Organisms and Populations.” *Current Opinion in Structural Biology* 42 (February): 31–40. https://doi.org/10.1016/j.sbi.2016.10.013.

Bloom, J. D., S. T. Labthavikul, C. R. Otey, and F. H. Arnold. 2006. “Protein Stability Promotes Evolvability.” *Proceedings of the National Academy of Sciences* 103 (15): 5869–74. https://doi.org/10.1073/pnas.0510098103.

Bloom, J. D., J. J. Silberg, C. O. Wilke, D. A. Drummond, C. Adami, and F. H. Arnold. 2005. “Thermodynamic Prediction of Protein Neutrality.” *Proceedings of the National Academy of Sciences* 102 (3): 606–11. https://doi.org/10.1073/pnas.0406744102.

Bloom, Jesse D., and Matthew J. Glassman. 2009. “Inferring Stabilizing Mutations from Protein Phylogenies: Application to Influenza Hemagglutinin.” Edited by Eugene I. Shakhnovich. *PLoS Computational Biology* 5 (4): e1000349. https://doi.org/10.1371/journal.pcbi.1000349.

Blount, Z. D., C. Z. Borland, and R. E. Lenski. 2008. “Historical Contingency and the Evolution of a Key Innovation in an Experimental Population of *Escherichia Coli*.” *Proceedings of the National Academy of Sciences* 105 (23): 7899–7906. https://doi.org/10.1073/pnas.0803151105.

Breen, Michael S., Carsten Kemena, Peter K. Vlasov, Cedric Notredame, and Fyodor A. Kondrashov. 2012. "Epistasis as the Primary Factor in Molecular Evolution." *Nature* 490 (7421): 535–38. https://doi.org/10.1038/nature11510.

Brockhurst, Michael A. 2007. "Population Bottlenecks Promote Cooperation in Bacterial Biofilms." Edited by Erik Svensson. *PLoS ONE* 2 (7): e634. https://doi.org/10.1371/journal.pone.0000634.

Brown, Nicholas G., Jeanine M. Pennington, Wanzhi Huang, Tulin Ayvaz, and Timothy Palzkill. 2010. "Multiple Global Suppressors of Protein Stability Defects Facilitate the Evolution of Extended-Spectrum TEM β-Lactamases." *Journal of Molecular Biology* 404 (5): 832–46. https://doi.org/10.1016/j.jmb.2010.10.008.

Cao, Yang, Daniel T. Gillespie, and Linda R. Petzold. 2007. "Adaptive Explicit-Implicit Tau-Leaping Method with Automatic Tau Selection." *The Journal of Chemical Physics* 126 (22): 224101. https://doi.org/10.1063/1.2745299.

Charlesworth, Brian. 2009. "Effective Population Size and Patterns of Molecular Evolution and Variation." *Nature Reviews Genetics* 10 (3): 195–205. https://doi.org/10.1038/nrg2526.

Chen, P., and E. I. Shakhnovich. 2009. "Lethal Mutagenesis in Viruses and Bacteria." *Genetics* 183 (2): 639–50. https://doi.org/10.1534/genetics.109.106492.

Chéron, Nicolas, Adrian W.R. Serohijos, Jeong-Mo Choi, and Eugene I. Shakhnovich. 2016. "Evolutionary Dynamics of Viral Escape under Antibodies Stress: A Biophysical Model: Evolutionary Dynamics of Viral Escape Under Antibodies Stress." *Protein Science* 25 (7): 1332–40. https://doi.org/10.1002/pro.2915.

Chevin, Luis-Miguel. 2011. "On Measuring Selection in Experimental Evolution." *Biology Letters* 7 (2): 210–13. https://doi.org/10.1098/rsbl.2010.0580.

Combe, Marine, and Rafael Sanjuán. 2014. "Variation in RNA Virus Mutation Rates across Host Cells." Edited by Adam Lauring. *PLoS Pathogens* 10 (1): e1003855. https://doi.org/10.1371/journal.ppat.1003855.

Cooksey, R, J Swenson, N Clark, E Gay, and C Thornsberry. 1990. "Patterns and Mechanisms of Beta-Lactam Resistance among Isolates of *Escherichia Coli* from Hospitals in the United States." *Antimicrobial Agents and Chemotherapy* 34 (5): 739–45. https://doi.org/10.1128/AAC.34.5.739.

Cowperthwaite, Matthew C., J. J. Bull, and Lauren Ancel Meyers. 2006. "From Bad to Good: Fitness Reversals and the Ascent of Deleterious Mutations." *PLoS Computational Biology* 2 (10): e141. https://doi.org/10.1371/journal.pcbi.0020141.

Crécy-Lagard, Valérie de. 2014. "Variations in Metabolic Pathways Create Challenges for Automated Metabolic Reconstructions: Examples from the Tetrahydrofolate Synthesis Pathway." *Computational and Structural Biotechnology Journal* 10 (16): 41–50. https://doi.org/10.1016/j.csbj.2014.05.008.

Dasmeh, Pouria, Éric Girard, and Adrian W. R. Serohijos. 2017. "Highly Expressed Genes Evolve under Strong Epistasis from a Proteome-Wide Scan in *E. Coli*." *Scientific Reports* 7 (1). https://doi.org/10.1038/s41598-017-16030-z.

Dasmeh, Pouria, and Adrian Serohijos. 2017. "Estimating The Contribution Of Folding Stability To Non-Specific Epistasis In Protein Evolution," March. https://doi.org/10.1101/122259.

Dasmeh, Pouria, Adrian W. R. Serohijos, Kasper P. Kepp, and Eugene I. Shakhnovich. 2013. "Positively Selected Sites in Cetacean Myoglobins Contribute to Protein Stability." Edited by Nir Ben-Tal. *PLoS Computational Biology* 9 (3): e1002929. https://doi.org/10.1371/journal.pcbi.1002929.

Dasmeh, Pouria, Adrian W.R. Serohijos, Kasper P. Kepp, and Eugene I. Shakhnovich. 2014. "The Influence of Selection for Protein Stability on DN/DS Estimations." *Genome Biology and Evolution* 6 (10): 2956–67. https://doi.org/10.1093/gbe/evu223.

DePristo, Mark A., Daniel M. Weinreich, and Daniel L. Hartl. 2005. "Missense Meanderings in Sequence Space: A Biophysical View of Protein Evolution." *Nature Reviews Genetics* 6 (9): 678–87. https://doi.org/10.1038/nrg1672.

Dieckmann, U., and R. Law. 1996. "The Dynamical Theory of Coevolution: A Derivation from Stochastic Ecological Processes." *Journal of Mathematical Biology* 34 (5–6): 579–612.

Drummond, D. A., J. D. Bloom, C. Adami, C. O. Wilke, and F. H. Arnold. 2005. "Why Highly Expressed Proteins Evolve Slowly." *Proceedings of the National Academy of Sciences* 102 (40): 14338–43. https://doi.org/10.1073/pnas.0504070102.

Edgar, Robert C. 2010. "Search and Clustering Orders of Magnitude Faster than BLAST." *Bioinformatics* 26 (19): 2460–61. https://doi.org/10.1093/bioinformatics/btq461.

Enne, V. I. 2010. "Reducing Antimicrobial Resistance in the Community by Restricting Prescribing: Can It Be Done?" *Journal of Antimicrobial Chemotherapy* 65 (2): 179–82. https://doi.org/10.1093/jac/dkp443.

Enright, M. C., D. A. Robinson, G. Randle, E. J. Feil, H. Grundmann, and B. G. Spratt. 2002. "The Evolutionary History of Methicillin-Resistant Staphylococcus Aureus (MRSA)." *Proceedings of the National Academy of Sciences* 99 (11): 7687–92. https://doi.org/10.1073/pnas.122108599.

Eyrewalker, A. 2006. "The Genomic Rate of Adaptive Evolution." *Trends in Ecology & Evolution* 21 (10): 569–75. https://doi.org/10.1016/j.tree.2006.06.015.

Eyre-Walker, Adam, and Peter D. Keightley. 2007. "The Distribution of Fitness Effects of New Mutations." *Nature Reviews Genetics* 8 (8): 610–18. https://doi.org/10.1038/nrg2146.

Firnberg, Elad, and Marc Ostermeier. 2012. "PFunkel: Efficient, Expansive, User-Defined Mutagenesis." Edited by D. Dafydd Jones. *PLoS ONE* 7 (12): e52031. https://doi.org/10.1371/journal.pone.0052031.

Fowler, Douglas M, Carlos L Araya, Sarel J Fleishman, Elizabeth H Kellogg, Jason J Stephany, David Baker, and Stanley Fields. 2010. "High-Resolution Mapping of Protein Sequence-Function Relationships." *Nature Methods* 7 (9): 741–46. https://doi.org/10.1038/nmeth.1492.

Fowler, Douglas M, and Stanley Fields. 2014. "Deep Mutational Scanning: A New Style of Protein Science." *Nature Methods* 11 (8): 801–7. https://doi.org/10.1038/nmeth.3027.

Fowler, Douglas M, Jason J Stephany, and Stanley Fields. 2014. "Measuring the Activity of Protein Variants on a Large Scale Using Deep Mutational Scanning." *Nature Protocols* 9 (9): 2267–84. https://doi.org/10.1038/nprot.2014.153.

Genolini, Christophe, Xavier Alacoque, Marianne Sentenac, and Catherine Arnaud. 2015. "**Kml** and **Kml3d** : *R* Packages to Cluster Longitudinal Data." *Journal of Statistical Software* 65 (4). https://doi.org/10.18637/jss.v065.i04.

Genolini, Christophe, René Ecochard, Mamoun Benghezal, Tarak Driss, Sandrine Andrieu, and Fabien Subtil. 2016. "KmlShape: An Efficient Method to Cluster Longitudinal Data (Time-Series) According to Their Shapes." Edited by Chun-Hsi Huang. *PLOS ONE* 11 (6): e0150738. https://doi.org/10.1371/journal.pone.0150738.

Genolini, Christophe, and Bruno Falissard. 2011. "Kml: A Package to Cluster Longitudinal Data." *Computer Methods and Programs in Biomedicine* 104 (3): e112–21. https://doi.org/10.1016/j.cmpb.2011.05.008.

Gillespie, Daniel T. 2001. "Approximate Accelerated Stochastic Simulation of Chemically Reacting Systems." *The Journal of Chemical Physics* 115 (4): 1716–33. https://doi.org/10.1063/1.1378322.

Goldstein, Richard A. 2011. "The Evolution and Evolutionary Consequences of Marginal Thermostability in Proteins." *Proteins: Structure, Function, and Bioinformatics* 79 (5): 1396–1407. https://doi.org/10.1002/prot.22964.

———. 2013. "Population Size Dependence of Fitness Effect Distribution and Substitution Rate Probed by Biophysical Model of Protein Thermostability." *Genome Biology and Evolution* 5 (9): 1584–93. https://doi.org/10.1093/gbe/evt110.

Good, Benjamin H., Michael J. McDonald, Jeffrey E. Barrick, Richard E. Lenski, and Michael M. Desai. 2017. "The Dynamics of Molecular Evolution over 60,000 Generations." *Nature* 551 (7678): 45–50. https://doi.org/10.1038/nature24287.

Gottesman, Bat Sheva, Yehuda Carmeli, Pnina Shitrit, and Michal Chowers. 2009. "Impact of Quinolone Restriction on Resistance Patterns of *Escherichia Coli* Isolated from Urine by Culture in a Community Setting." *Clinical Infectious Diseases* 49 (6): 869–75. https://doi.org/10.1086/605530.

Goyal, S., D. J. Balick, E. R. Jerison, R. A. Neher, B. I. Shraiman, and M. M. Desai. 2012. "Dynamic Mutation-Selection Balance as an Evolutionary Attractor." *Genetics* 191 (4): 1309–19. https://doi.org/10.1534/genetics.112.141291.

Guo, Beining, Kamilia Abdelraouf, Kimberly R. Ledesma, Michael Nikolaou, and Vincent H. Tam. 2012. "Predicting Bacterial Fitness Cost Associated with Drug Resistance." *Journal of Antimicrobial Chemotherapy* 67 (4): 928–32. https://doi.org/10.1093/jac/dkr560.

Gupta, Aditi, and Christoph Adami. 2016. "Strong Selection Significantly Increases Epistatic Interactions in the Long-Term Evolution of a Protein." Edited by Jianzhi Zhang. *PLOS Genetics* 12 (3): e1005960. https://doi.org/10.1371/journal.pgen.1005960.

Gutiérrez, Serafín, Yannis Michalakis, and Stéphane Blanc. 2012. "Virus Population Bottlenecks during Within-Host Progression and Host-to-Host Transmission." *Current Opinion in Virology* 2 (5): 546–55. https://doi.org/10.1016/j.coviro.2012.08.001.

Hall, B. G. 2002. "Predicting Evolution by In Vitro Evolution Requires Determining Evolutionary Pathways." *Antimicrobial Agents and Chemotherapy* 46 (9): 3035–38. https://doi.org/10.1128/AAC.46.9.3035-3038.2002.

Hao, Jihua, Adrian W. R. Serohijos, Gail Newton, Gina Tassone, Zuncai Wang, Dennis C. Sgroi, Nikolay V. Dokholyan, and James P. Basilion. 2008. "Identification and Rational Redesign of Peptide Ligands to CRIP1, A Novel Biomarker for Cancers." Edited by Eugene I. Shakhnovich. *PLoS Computational Biology* 4 (8): e1000138. https://doi.org/10.1371/journal.pcbi.1000138.

Harms, Michael J., and Joseph W. Thornton. 2013. "Evolutionary Biochemistry: Revealing the Historical and Physical Causes of Protein Properties." *Nature Reviews Genetics* 14 (8): 559–71. https://doi.org/10.1038/nrg3540.

Harper, Marc, and Dashiell Fryer. 2016. "Stationary Stability for Evolutionary Dynamics in Finite Populations." *Entropy* 18 (9): 316. https://doi.org/10.3390/e18090316.

Hartl, Daniel L. 2014. "What Can We Learn from Fitness Landscapes?" *Current Opinion in Microbiology* 21 (October): 51–57. https://doi.org/10.1016/j.mib.2014.08.001.

Hartl, Daniel L., and Andrew G. Clark. 2007. *Principles of Population Genetics*. 4th ed. Sunderland, Mass: Sinauer Associates.

Higgins, Christopher F. 2007. "Multiple Molecular Mechanisms for Multidrug Resistance Transporters." *Nature* 446 (7137): 749–57. https://doi.org/10.1038/nature05630.

"How Populations Grow: The Exponential and Logistic Equations | Learn Science at Scitable." n.d. Accessed May 7, 2018. https://www.nature.com/scitable/knowledge/library/how-populations-grow-the-exponential-and-logistic-13240157.

Imtiaz, U, E K Manavathu, S Mobashery, and S A Lerner. 1994. "Reversal of Clavulanate Resistance Conferred by a Ser-244 Mutant of TEM-1 Beta-Lactamase as a Result of a Second Mutation (Arg to Ser at Position 164) That Enhances Activity against Ceftazidime." *Antimicrobial Agents and Chemotherapy* 38 (5): 1134–39. https://doi.org/10.1128/AAC.38.5.1134.

Johnsen, Pål J, Jeffrey P Townsend, Thomas Bøhn, Gunnar S Simonsen, Arnfinn Sundsfjord, and Kaare M Nielsen. 2009. "Factors Affecting the Reversal of Antimicrobial-Drug Resistance." *The Lancet Infectious Diseases* 9 (6): 357–64. https://doi.org/10.1016/S1473-3099(09)70105-7.

Johnson, Philip. 2016. “Adaptivetau: Efficient Stochastic Simulations in R,” 10.

Kimura, Motoo. 1968. “Evolutionary Rate at the Molecular Level.” *Nature* 217 (5129): 624–26. https://doi.org/10.1038/217624a0.

Kliman, Richard, Bob Sheehy, and Joanna Schultz. 2008. “Genetic Drift and Effective Population Size.” *Nature Education* 1 (3): 3.

Knöppel, Anna, Joakim Näsvall, and Dan I. Andersson. 2017. “Evolution of Antibiotic Resistance without Antibiotic Exposure.” *Antimicrobial Agents and Chemotherapy* 61 (11). https://doi.org/10.1128/AAC.01495-17.

Kong, Kok-Fai, Lisa Schneper, and Kalai Mathee. 2010. “Beta-Lactam Antibiotics: From Antibiosis to Resistance and Bacteriology.” *APMIS: Acta Pathologica, Microbiologica, et Immunologica Scandinavica* 118 (1): 1–36. https://doi.org/10.1111/j.1600-0463.2009.02563.x.

Koolman, J., and P. Karlson. 1975. “Ecdysone Oxidase, an Enzyme from the Blowfly Calliphora Erythrocephala (Meigen).” *Hoppe-Seyler’s Zeitschrift Fur Physiologische Chemie* 356 (7): 1131–38.

Kowalsky, Caitlin A., Matthew S. Faber, Aritro Nath, Hailey E. Dann, Vince W. Kelly, Li Liu, Purva Shanker, et al. 2015. “Rapid Fine Conformational Epitope Mapping Using Comprehensive Mutagenesis and Deep Sequencing.” *Journal of Biological Chemistry* 290 (44): 26457–70. https://doi.org/10.1074/jbc.M115.676635.

Kowalsky, Caitlin A., Justin R. Klesmith, James A. Stapleton, Vince Kelly, Nolan Reichkitzer, and Timothy A. Whitehead. 2015. “High-Resolution Sequence-Function Mapping of Full-Length Proteins.” Edited by Gideon Schreiber. *PLOS ONE* 10 (3): e0118193. https://doi.org/10.1371/journal.pone.0118193.

Lee, Chang-Ro, Ill Cho, Byeong Jeong, and Sang Lee. 2013. “Strategies to Minimize Antibiotic Resistance.” *International Journal of Environmental Research and Public Health* 10 (9): 4274–4305. https://doi.org/10.3390/ijerph10094274.

Levin, B. R., V. Perrot, and N. Walker. 2000. “Compensatory Mutations, Antibiotic Resistance and the Population Genetics of Adaptive Evolution in Bacteria.” *Genetics* 154 (3): 985–97.

Liberles, David A., Sarah A. Teichmann, Ivet Bahar, Ugo Bastolla, Jesse Bloom, Erich Bornberg-Bauer, Lucy J. Colwell, et al. 2012. “The Interface of Protein Structure, Protein Biophysics, and Molecular Evolution.” *Protein Science* 21 (6): 769–85. https://doi.org/10.1002/pro.2071.

Lipinski, Kamil A., Louise J. Barber, Matthew N. Davies, Matthew Ashenden, Andrea Sottoriva, and Marco Gerlinger. 2016. "Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine." *Trends in Cancer* 2 (1): 49–63. https://doi.org/10.1016/j.trecan.2015.11.003.

Lipsitch, M., C. T. Bergstrom, and B. R. Levin. 2000. "The Epidemiology of Antibiotic Resistance in Hospitals: Paradoxes and Prescriptions." *Proceedings of the National Academy of Sciences of the United States of America* 97 (4): 1938–43.

Lobkovsky, Alexander E., Yuri I. Wolf, and Eugene V. Koonin. 2011. "Predictability of Evolutionary Trajectories in Fitness Landscapes." Edited by Eugene I. Shakhnovich. *PLoS Computational Biology* 7 (12): e1002302. https://doi.org/10.1371/journal.pcbi.1002302.

Loewe, L., and W. G. Hill. 2010. "The Population Genetics of Mutations: Good, Bad and Indifferent." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1544): 1153–67. https://doi.org/10.1098/rstb.2009.0317.

Ma, Ling, Yoshikazu Ishii, Keizo Yamaguchi, and Hirosuke Matsuo. 2002. "Characterization of Cefotaxime-Resistant *Escherichia Coli* Isolates from a Nosocomial Outbreak at Three Geriatric Hospitals." *Journal of Infection and Chemotherapy* 8 (2): 155–62. https://doi.org/10.1007/s101560200027.

Mackay, Trudy F. C. 2014. "Epistasis and Quantitative Traits: Using Model Organisms to Study Gene–Gene Interactions." *Nature Reviews Genetics* 15 (1): 22–33. https://doi.org/10.1038/nrg3627.

Maharjan, Ram, and Thomas Ferenci. 2017. "The Fitness Costs and Benefits of Antibiotic Resistance in Drug-Free Microenvironments Encountered in the Human Body: Fitness Costs of Antibiotic Resistance." *Environmental Microbiology Reports* 9 (5): 635–41. https://doi.org/10.1111/1758-2229.12564.

Maisnier-Patin, Sophie, and Dan I Andersson. 2004. "Adaptation to the Deleterious Effects of Antimicrobial Drug Resistance Mutations by Compensatory Evolution." *Research in Microbiology* 155 (5): 360–69. https://doi.org/10.1016/j.resmic.2004.01.019.

Martinez, J. L., and F. Baquero. 2000. "Mutation Frequencies and Antibiotic Resistance." *Antimicrobial Agents and Chemotherapy* 44 (7): 1771–77. https://doi.org/10.1128/AAC.44.7.1771-1777.2000.

Martínez, José L., Fernando Baquero, and Dan I. Andersson. 2007. "Predicting Antibiotic Resistance." *Nature Reviews Microbiology* 5 (12): 958–65. https://doi.org/10.1038/nrmicro1796.

Matagne, A., J. Lamotte-Brasseur, and J. M. Frère. 1998. "Catalytic Properties of Class A Beta-Lactamases: Efficiency and Diversity." *The Biochemical Journal* 330 (Pt 2) (March): 581–98.

McCandlish, David M., Etienne Rajon, Premal Shah, Yang Ding, and Joshua B. Plotkin. 2013. "The Role of Epistasis in Protein Evolution." *Nature* 497 (7451): E1–2. https://doi.org/10.1038/nature12219.

McLaughlin Jr, Richard N., Frank J. Poelwijk, Arjun Raman, Walraj S. Gosal, and Rama Ranganathan. 2012. "The Spatial Architecture of Protein Function and Adaptation." *Nature* 491 (7422): 138–42. https://doi.org/10.1038/nature11500.

Melnyk, Anita H., Alex Wong, and Rees Kassen. 2015. "The Fitness Costs of Antibiotic Resistance Mutations." *Evolutionary Applications* 8 (3): 273–83. https://doi.org/10.1111/eva.12196.

Meyer, Austin G., and Claus O. Wilke. 2013. "Integrating Sequence Variation and Protein Structure to Identify Sites under Selection." *Molecular Biology and Evolution* 30 (1): 36–44. https://doi.org/10.1093/molbev/mss217.

Moran, N. A. 1996. "Accelerated Evolution and Muller's Rachet in Endosymbiotic Bacteria." *Proceedings of the National Academy of Sciences* 93 (7): 2873–78. https://doi.org/10.1073/pnas.93.7.2873.

Muirhead, Christina A., and John Wakeley. 2009. "Modeling Multiallelic Selection Using a Moran Model." *Genetics* 182 (4): 1141–57. https://doi.org/10.1534/genetics.108.089474.

Myllykallio, Hannu, Damien Leduc, Jonathan Filee, and Ursula Liebl. 2003. "Life without Dihydrofolate Reductase FolA." *Trends in Microbiology* 11 (5): 220–23. https://doi.org/10.1016/S0966-842X(03)00101-X.

Nowak, M. A. 2006. *Evolutionary Dynamics: Exploring the Equations of Life*. Cambridge, Mass: Belknap Press of Harvard University Press.

Olivares, Jorge, Carolina Álvarez-Ortega, and José Luis Martinez. 2014. "Metabolic Compensation of Fitness Costs Associated with Overexpression of the Multidrug Efflux Pump MexEF-OprN in Pseudomonas aeruginosa." *Antimicrobial Agents and Chemotherapy* 58 (7): 3904–13. https://doi.org/10.1128/AAC.00121-14.

Olofsson, S. K., P. Geli, D. I. Andersson, and O. Cars. 2005. "Pharmacodynamic Model To Describe the Concentration-Dependent Selection of Cefotaxime-Resistant *Escherichia Coli*." *Antimicrobial Agents and Chemotherapy* 49 (12): 5081–91. https://doi.org/10.1128/AAC.49.12.5081-5091.2005.

Orr, H. Allen. 2003. "The Distribution of Fitness Effects among Beneficial Mutations." *Genetics* 163 (4): 1519–26.

———. 2009. "Fitness and Its Role in Evolutionary Genetics." *Nature Reviews Genetics* 10 (8): 531–39. https://doi.org/10.1038/nrg2603.

Osmond, M. M., and C. de Mazancourt. 2012. "How Competition Affects Evolutionary Rescue." *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1610): 20120085–20120085. https://doi.org/10.1098/rstb.2012.0085.

Palmer, Adam C., and Roy Kishony. 2013. "Understanding, Predicting and Manipulating the Genotypic Evolution of Antibiotic Resistance." *Nature Reviews Genetics* 14 (4): 243–48. https://doi.org/10.1038/nrg3351.

Paterson, Iona K., Andy Hoyle, Gabriela Ochoa, Craig Baker-Austin, and Nick G. H. Taylor. 2016. "Optimising Antibiotic Usage to Treat Bacterial Infections." *Scientific Reports* 6 (1). https://doi.org/10.1038/srep37853.

Perez-Figueroa, A., A. Caballero, A. Garcia-Dorado, and C. Lopez-Fanjul. 2009. "The Action of Purifying Selection, Mutation and Drift on Fitness Epistatic Systems." *Genetics* 183 (1): 299–313. https://doi.org/10.1534/genetics.109.104893.

Perron, Gabriel G., Sergey Kryazhimskiy, Daniel P. Rice, and Angus Buckling. 2012. "Multidrug Therapy and Evolution of Antibiotic Resistance: When Order Matters." *Applied and Environmental Microbiology* 78 (17): 6137–42. https://doi.org/10.1128/AEM.01078-12.

Phillips, Patrick C. 2008. "Epistasis--the Essential Role of Gene Interactions in the Structure and Evolution of Genetic Systems." *Nature Reviews. Genetics* 9 (11): 855–67. https://doi.org/10.1038/nrg2452.

Piganeau, G., and A. Eyre-Walker. 2003. "Estimating the Distribution of Fitness Effects from DNA Sequence Data: Implications for the Molecular Clock." *Proceedings of the National Academy of Sciences* 100 (18): 10335–40. https://doi.org/10.1073/pnas.1833064100.

Poelwijk, Frank J., Daniel J. Kiviet, Daniel M. Weinreich, and Sander J. Tans. 2007. "Empirical Fitness Landscapes Reveal Accessible Evolutionary Paths." *Nature* 445 (7126): 383–86. https://doi.org/10.1038/nature05451.

R Core Team. 2013. "R: A Language and Environment for Statistical Computing." *R Foundation for Statistical Computing*. http://www.R-project.org/.

Robicsek, Ari, Jacob Strahilevitz, George A Jacoby, Mark Macielag, Darren Abbanat, Chi Hye Park, Karen Bush, and David C Hooper. 2006. "Fluoroquinolone-Modifying Enzyme: A New Adaptation of a Common Aminoglycoside Acetyltransferase." *Nature Medicine* 12 (1): 83–88. https://doi.org/10.1038/nm1347.

Rodrigues, João V., Shimon Bershtein, Anna Li, Elena R. Lozovsky, Daniel L. Hartl, and Eugene I. Shakhnovich. 2016. "Biophysical Principles Predict Fitness Landscapes of Drug Resistance." *Proceedings of the National Academy of Sciences* 113 (11): E1470–78. https://doi.org/10.1073/pnas.1601441113.

Rodríguez-Verdugo, Alejandra, Olivier Tenaillon, and Brandon S. Gaut. 2016. "First-Step Mutations during Adaptation Restore the Expression of Hundreds of Genes." *Molecular Biology and Evolution* 33 (1): 25–39. https://doi.org/10.1093/molbev/msv228.

Roux, Damien, Olga Danilchanka, Thomas Guillard, Vincent Cattoir, Hugues Aschard, Yang Fu, Francois Angoulvant, et al. 2015. "Fitness Cost of Antibiotic Susceptibility during Bacterial Infection." *Science Translational Medicine* 7 (297): 297ra114-297ra114. https://doi.org/10.1126/scitranslmed.aab1621.

Rouzine, I. M., A. Rodrigo, and J. M. Coffin. 2001. "Transition between Stochastic Evolution and Deterministic Evolution in the Presence of Selection: General Theory and Application to Virology." *Microbiology and Molecular Biology Reviews* 65 (1): 151–85. https://doi.org/10.1128/MMBR.65.1.151-185.2001.

Rubin, Alan F., Hannah Gelman, Nathan Lucas, Sandra M. Bajjalieh, Anthony T. Papenfuss, Terence P. Speed, and Douglas M. Fowler. 2017. "A Statistical Framework for Analyzing Deep Mutational Scanning Data." *Genome Biology* 18 (1). https://doi.org/10.1186/s13059-017-1272-5.

Sailer, Zachary R., and Michael J. Harms. 2017. "High-Order Epistasis Shapes Evolutionary Trajectories." Edited by Joachim Krug. *PLOS Computational Biology* 13 (5): e1005541. https://doi.org/10.1371/journal.pcbi.1005541.

Salverda, Merijn L. M., Jeroen Koomen, Bertha Koopmanschap, Mark P. Zwart, and J. Arjan G. M. de Visser. 2017. "Adaptive Benefits from Small Mutation Supplies in an Antibiotic Resistance Enzyme." *Proceedings of the National Academy of Sciences* 114 (48): 12773–78. https://doi.org/10.1073/pnas.1712999114.

Salverda, Merijn L.M., J. Arjan G.M. De Visser, and Miriam Barlow. 2010. "Natural Evolution of TEM-1 β-Lactamase: Experimental Reconstruction and Clinical Relevance." *FEMS Microbiology Reviews* 34 (6): 1015–36. https://doi.org/10.1111/j.1574-6976.2010.00222.x.

Sanjuan, R., A. Moya, and S. F. Elena. 2004. "The Distribution of Fitness Effects Caused by Single-Nucleotide Substitutions in an RNA Virus." *Proceedings of the National Academy of Sciences* 101 (22): 8396–8401. https://doi.org/10.1073/pnas.0400146101.

Schlosser, Gerhard, and Günter P. Wagner. 2008. "A Simple Model of Co-Evolutionary Dynamics Caused by Epistatic Selection." *Journal of Theoretical Biology* 250 (1): 48–65. https://doi.org/10.1016/j.jtbi.2007.08.033.

Schrag, S. J., V. Perrot, and B. R. Levin. 1997. "Adaptation to the Fitness Costs of Antibiotic Resistance in *Escherichia Coli*." *Proceedings of the Royal Society B: Biological Sciences* 264 (1386): 1287–91. https://doi.org/10.1098/rspb.1997.0178.

Schulz zur Wiesch, P., J. Engelstadter, and S. Bonhoeffer. 2010. "Compensation of Fitness Costs and Reversibility of Antibiotic Resistance Mutations." *Antimicrobial Agents and Chemotherapy* 54 (5): 2085–95. https://doi.org/10.1128/AAC.01460-09.

Serohijos, Adrian W.R., and Eugene I. Shakhnovich. 2014. "Contribution of Selection for Protein Folding Stability in Shaping the Patterns of Polymorphisms in Coding Regions." *Molecular Biology and Evolution* 31 (1): 165–76. https://doi.org/10.1093/molbev/mst189.

Serohijos, Adrian WR, and Eugene I Shakhnovich. 2014. "Merging Molecular Mechanism and Evolution: Theory and Computation at the Interface of Biophysics and Evolutionary Population Genetics." *Current Opinion in Structural Biology* 26 (June): 84–91. https://doi.org/10.1016/j.sbi.2014.05.005.

Serohijos, Adrian W.R., S. Y. Ryan Lee, and Eugene I. Shakhnovich. 2013. "Highly Abundant Proteins Favor More Stable 3D Structures in Yeast." *Biophysical Journal* 104 (3): L1–3. https://doi.org/10.1016/j.bpj.2012.11.3838.

Serohijos, Adrian W.R., Zilvinas Rimas, and Eugene I. Shakhnovich. 2012. "Protein Biophysics Explains Why Highly Abundant Proteins Evolve Slowly." *Cell Reports* 2 (2): 249–56. https://doi.org/10.1016/j.celrep.2012.06.022.

Shah, Premal, David M. McCandlish, and Joshua B. Plotkin. 2015. "Contingency and Entrenchment in Protein Evolution under Purifying Selection." *Proceedings of the National Academy of Sciences* 112 (25): E3226–35. https://doi.org/10.1073/pnas.1412933112.

Shaikh, Sibhghatulla, Jamale Fatima, Shazi Shakil, Syed Mohd. Danish Rizvi, and Mohammad Amjad Kamal. 2015. "Antibiotic Resistance and Extended Spectrum Beta-Lactamases: Types, Epidemiology and Treatment." *Saudi Journal of Biological Sciences* 22 (1): 90–101. https://doi.org/10.1016/j.sjbs.2014.08.002.

Shao, H., L. C. Burrage, D. S. Sinasac, A. E. Hill, S. R. Ernest, W. O'Brien, H.-W. Courtland, et al. 2008. "Genetic Architecture of Complex Traits: Large Phenotypic Effects and Pervasive Epistasis." *Proceedings of the National Academy of Sciences* 105 (50): 19910–14. https://doi.org/10.1073/pnas.0810388105.

Sikosek, T., and H. S. Chan. 2014. "Biophysics of Protein Evolution and Evolutionary Protein Biophysics." *Journal of The Royal Society Interface* 11 (100): 20140419–20140419. https://doi.org/10.1098/rsif.2014.0419.

Silander, Olin K, Olivier Tenaillon, and Lin Chao. 2007. "Understanding the Evolutionary Fate of Finite Populations: The Dynamics of Mutational Effects." Edited by Nick H Barton. *PLoS Biology* 5 (4): e94. https://doi.org/10.1371/journal.pbio.0050094.

Stiffler, Michael A., Doeke R. Hekstra, and Rama Ranganathan. 2015. "Evolvability as a Function of Purifying Selection in TEM-1 β-Lactamase." *Cell* 160 (5): 882–92. https://doi.org/10.1016/j.cell.2015.01.035.

Sundqvist, M., P. Geli, D. I. Andersson, M. Sjolund-Karlsson, A. Runehagen, H. Cars, K. Abelson-Storby, O. Cars, and G. Kahlmeter. 2010. "Little Evidence for Reversibility of Trimethoprim Resistance after a Drastic Reduction in Trimethoprim Use." *Journal of Antimicrobial Chemotherapy* 65 (2): 350–60. https://doi.org/10.1093/jac/dkp387.

Sundqvist, Martin. 2014. "Reversibility of Antibiotic Resistance." *Upsala Journal of Medical Sciences* 119 (2): 142–48. https://doi.org/10.3109/03009734.2014.903323.

Szpara, M. L., L. Parsons, and L. W. Enquist. 2010. "Sequence Variability in Clinical and Laboratory Isolates of Herpes Simplex Virus 1 Reveals New Mutations." *Journal of Virology* 84 (10): 5303–13. https://doi.org/10.1128/JVI.00312-10.

Takesue, Y., K. Nakajima, K. Ichiki, M. Ishihara, Y. Wada, Y. Takahashi, T. Tsuchida, and H. Ikeuchi. 2010. "Impact of a Hospital-Wide Programme of Heterogeneous Antibiotic Use on the Development of Antibiotic-Resistant Gram-Negative Bacteria." *Journal of Hospital Infection* 75 (1): 28–32. https://doi.org/10.1016/j.jhin.2009.11.022.

Toprak, Erdal, Adrian Veres, Jean-Baptiste Michel, Remy Chait, Daniel L Hartl, and Roy Kishony. 2012. "Evolutionary Paths to Antibiotic Resistance under Dynamically Sustained Drug Selection." *Nature Genetics* 44 (1): 101–5. https://doi.org/10.1038/ng.1034.

Torella, Joseph Peter, Remy Chait, and Roy Kishony. 2010. "Optimal Drug Synergy in Antimicrobial Treatments." Edited by Philip E. Bourne. *PLoS Computational Biology* 6 (6): e1000796. https://doi.org/10.1371/journal.pcbi.1000796.

Ueda, Masahiko, Nobuto Takeuchi, and Kunihiko Kaneko. 2017. "Stronger Selection Can Slow down Evolution Driven by Recombination on a Smooth Fitness Landscape." Edited by Roeland M. H. Merks. *PLOS ONE* 12 (8): e0183120. https://doi.org/10.1371/journal.pone.0183120.

Ventola, C. Lee. 2015. "The Antibiotic Resistance Crisis: Part 1: Causes and Threats." *P & T: A Peer-Reviewed Journal for Formulary Management* 40 (4): 277–83.

Vihinen, Mauno. 2012. "How to Evaluate Performance of Prediction Methods? Measures and Their Interpretation in Variation Effect Analysis." *BMC Genomics* 13 (Suppl 4): S2. https://doi.org/10.1186/1471-2164-13-S4-S2.

Visser, J. Arjan G.M. de, and Joachim Krug. 2014. "Empirical Fitness Landscapes and the Predictability of Evolution." *Nature Reviews Genetics* 15 (7): 480–90. https://doi.org/10.1038/nrg3744.

Vogwill, Tom, and R. Craig MacLean. 2015. "The Genetic Basis of the Fitness Costs of Antimicrobial Resistance: A Meta-Analysis Approach." *Evolutionary Applications* 8 (3): 284–95. https://doi.org/10.1111/eva.12202.

Wakeley, John. 2009. *Coalescent Theory: An Introduction*. Greenwood Village, Colo: Roberts & Co. Publishers.

Wang, Ping, Qi Wang, Yonghong Yang, James K. Coward, Alexis Nzila, Paul F.G. Sims, and John E. Hyde. 2010. "Characterisation of the Bifunctional Dihydrofolate Synthase–Folylpolyglutamate Synthase from Plasmodium Falciparum; a Potential Novel Target for Antimalarial Antifolate Inhibition." *Molecular and Biochemical Parasitology* 172 (1): 41–51. https://doi.org/10.1016/j.molbiopara.2010.03.012.

Webber, Mark A., Rebekah N. Whitehead, Manuella Mount, Nick J. Loman, Mark J. Pallen, and Laura J. V. Piddock. 2015. "Parallel Evolutionary Pathways to Antibiotic Resistance Selected by Biocide Exposure." *The Journal of Antimicrobial Chemotherapy* 70 (8): 2241–48. https://doi.org/10.1093/jac/dkv109.

Weinreich, D. M. 2006. "Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins." *Science* 312 (5770): 111–14. https://doi.org/10.1126/science.1123539.

Wilke, Claus O. 2012. "Bringing Molecules Back into Molecular Evolution." Edited by Sergei L. Kosakovsky Pond. *PLoS Computational Biology* 8 (6): e1002572. https://doi.org/10.1371/journal.pcbi.1002572.

Wrenbeck, Emily E, Matthew S Faber, and Timothy A Whitehead. 2017. "Deep Sequencing Methods for Protein Engineering and Design." *Current Opinion in Structural Biology* 45 (August): 36–44. https://doi.org/10.1016/j.sbi.2016.11.001.

Wrenbeck, Emily E, Justin R Klesmith, James A Stapleton, Adebola Adeniran, Keith E J Tyo, and Timothy A Whitehead. 2016. "Plasmid-Based One-Pot Saturation Mutagenesis." *Nature Methods* 13 (11): 928–30. https://doi.org/10.1038/nmeth.4029.

Wright, Sewall. 1932. "The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution." *Proceedings of the Sixth International Congress of Genetics* 1: 356–66.

Wylie, C. S., and E. I. Shakhnovich. 2011. "A Biophysical Protein Folding Model Accounts for Most Mutational Fitness Effects in Viruses." *Proceedings of the National Academy of Sciences* 108 (24): 9916–21. https://doi.org/10.1073/pnas.1017572108.

Yang, Jian-Rong, Shi-Mei Zhuang, and Jianzhi Zhang. 2010. "Impact of Translational Error-Induced and Error-Free Misfolding on the Rate of Protein Evolution." *Molecular Systems Biology* 6 (October). https://doi.org/10.1038/msb.2010.78.

Yin, Shuangye, Feng Ding, and Nikolay V. Dokholyan. 2007. "Modeling Backbone Flexibility Improves Protein Stability Estimation." *Structure* 15 (12): 1567–76. https://doi.org/10.1016/j.str.2007.09.024.

Yun, M.-K., Y. Wu, Z. Li, Y. Zhao, M. B. Waddell, A. M. Ferreira, R. E. Lee, D. Bashford, and S. W. White. 2012. “Catalysis and Sulfa Drug Resistance in Dihydropteroate Synthase.” *Science* 335 (6072): 1110–14. https://doi.org/10.1126/science.1214641.

Zeldovich, K. B., P. Chen, and E. I. Shakhnovich. 2007. “Protein Stability Imposes Limits on Organism Complexity and Speed of Molecular Evolution.” *Proceedings of the National Academy of Sciences* 104 (41): 16152–57. https://doi.org/10.1073/pnas.0705366104.

Zhang, Huidan, Shelley K. Cockrell, Abimbola O. Kolawole, Assaf Rotem, Adrian W. R. Serohijos, Connie B. Chang, Ye Tao, et al. 2015. “Isolation and Analysis of Rare Norovirus Recombinants from Coinfected Mice Using Drop-Based Microfluidics.” Edited by S. R. Ross. *Journal of Virology* 89 (15): 7722–34. https://doi.org/10.1128/JVI.01137-15.

Appendix 1: TEM-1 WT Recovery Protocol

Site-directed mutagenesis to recover WT pSALECTNK-TEM1 using the QuikChange kit

0. DNA Template preparation.

Use overnight culture to determine the concentration of plasmid DNA (after mini-prep).

1. Phosphorylation of oligonucleotides.

Primers must be phosphorylated at the 5' end to eliminate the need for a separate phosphorylation step before direct ligation. Do the phosphorylation twice: once for the forward primer and once for the reverse primer.

1. Prepare the following reaction mixture in a centrifuge tube:
    a. 250 pmol oligonucleotide of **forward nucleotide** (29bp) -> Need 4.8µg. The concentration of primer is at 996ng/µL, use **4.8µL** to get 4.8µg (4800ng).
    b. 5uL 10X reaction buffer A for T4 polynucleotide kinase
    c. 4uL 10mM ATP
    d. 2uL T4 Polynucleotide Kinase 10U/µL
    e. NF-H2O to a final volume of 50µL

2. Prepare the following reaction mixture in another centrifuge tube:
    a. 250 pmol oligonucleotide of **MIDDLE reverse nucleotide** (20bp) -> 3.3µg. The concentration of primer is at 4933ng/µL, use **6.7µL** to get 3.3ug (3300ng).
    b. 5uL 10X reaction buffer A for T4 polynucleotide kinase
    c. 4uL 10mM ATP
    d. 2uL T4 Polynucleotide Kinase 10U/µL
    e. NF-H2O to a final volume of 50µL

3. Incubate reaction at 37°C for 30 minutes
4. Inactivate the T4 Polynucleotide kinase at 75°C for 10 minutes. **Alternatively, PNK reaction can be purified using a spin column kit.**
5. The reaction products can be stored at -20°C or added directly to the mutagenesis reaction. Use 5µL of 5'-phosphorylated oligonucleotide (5µM) for mutagenesis reaction.

2. PCR

To add (up to 50uL)	Component	Final Conc
10µL	5X Phusion HF Buffer	1X
1µL	10mM dNTPs	200µM each
5µL	Phosphorylated Forward Primer Phosphorylated D179G correction primer	0.5µM
5µL	Phosphorylated Reverse Primer MIDDLE reverse sequencing	0.5µM
XXXXX µL	Template DNA	
0.5µL	Phusion Hot Start DNA Polymerase	0.02U/µL
2.5µL	5% DMSO	
16µL	NF-H2O	

3. Cycling instructions for the mutagenesis reaction

Forward sequence information from Thermo-Fisher Tm calculator:

- Length: 29bp
- MW: 8904.9 g/mol
- Tm(°C): 77.0

Reverse sequence information from Thermo-Fisher Tm calculator:

- Length: 20bp
- MW: 6028.0 g/mol
- Tm(°C): 61.2

Suggested annealing temperature (°C): 68.0

Cycle step	Temp.	Time	Number of cycles
Initial denaturation	98°C	30s	1
Denaturation Annealing Extension	98°C 68°C 72°C	10s 25s 25s/kb -> 100s (since 4kb)	25
Final extension	72°C 4°C	7min hold	1

4. DpnI digestion of parental plasmid DNA

After PCR, add directly to the mutagenesis reaction:

- 1 uL of FastDigest DpnI enzyme directly to the mutagenesis reaction
- Incubate at 37°C for 15 minutes.

Determine the concentration of the PCR product -> ng/µL

5. Ligation

The PCR product is circularized using T4 DNA Ligase in a minute reaction. Transformation efficiency starts to decrease after 2 hours and is reduced by up to 75% if the reaction is allowed to proceed overnight at 25°C.

Prepare 10uL of the ligation mix

1. Take 10-20ng of PCR product from the mutagenesis reaction after DpnI digestion. This usually equals 1-5µL. **Use the concentration determined from step 4.** Do not use more than 5µL of PCR reaction mix for the 10uL ligation reaction.
2. Add 2µL of 5X Rapid Ligation Buffer
3. Adjust the reaction volume to 9.5µL of NF-H2O and mix
4. Add 0.5 µL of T4 DNA Ligase and mix thoroughly
5. Centrifuge briefly and incubate at room temperature (25°C) for 5 minutes
6. Chill on ice, then transform or store at -20°C

6. Transformation

Any standard *E. coli* strain that is suitable for DNA cloning can be used as a transformation host. Use electroporation. **It is recommended to use 1µL of the purified ligation mix per 50µL of electrocompetent *E. coli* cells.**

Follow the protocol for XL-1 blue transformation, use 1µL of XL-blue.

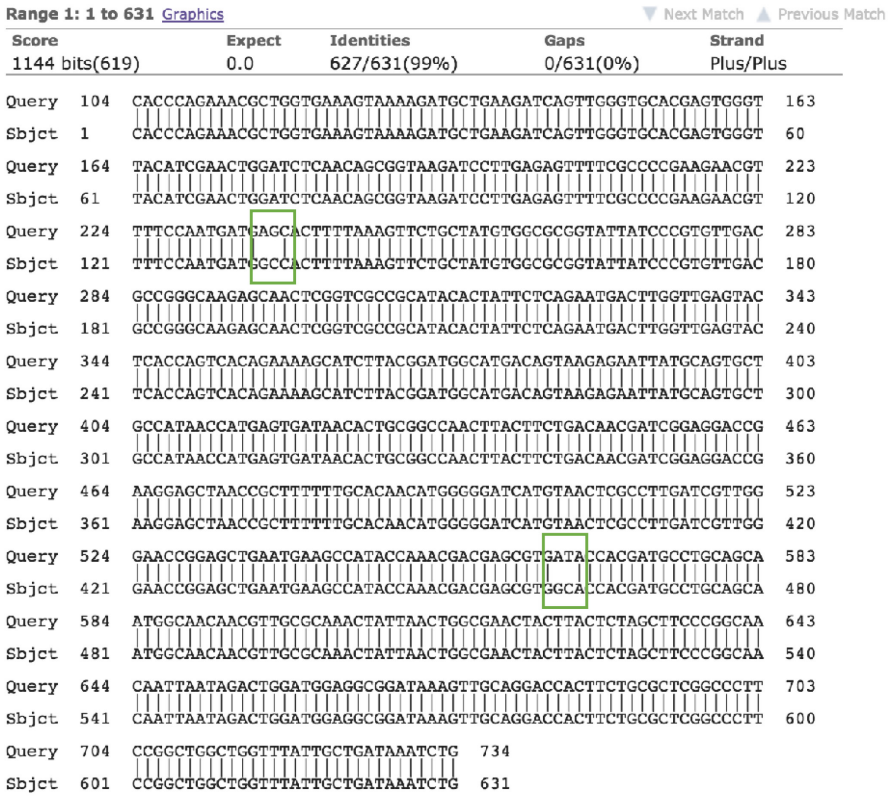

Figure 28. Alignment between the recovered TEM-1 WT strain and the p-salectnk TEM-1 plasmid.

Appendix 2: Nicking mutagenesis protocol for TEM-1

Mutagenesis Library Generation with Strain 127 (TEM-1 corrected WT) for Tile 1

i. Make a mixture of NNN/NNK mutagenic oligos at a final concentration of 10µM

- Use residues 0 to 66 (25 to 89), a total of 67 residues.
- Prepare a clean 96 well-plate
- Transfer 90µL NFH2O + 10µL of stock primer (C=100µM) for a final volume of 100µL of each primer at 10µM
- Transfer 10µL from all columns of the 96 plate to the same column of a new plate.
- Transfer all volume from each well of the combined column to an Eppendorf. This corresponds to the 10µM mutagenic oligo mixture.

1. Phosphorylate oligos

1. Make a mixture of oligos as described above.
2. Into a PCR tube, add:
    a. 20µL 10µM mutagenic oligo mixture (for tile 1, 2, 3, 4, and correction = 5 tubes)
    b. 2.4µL T4 PNK buffer
    c. 1µL 10mM ATP
    d. 1µL T4 PNK (10U/µL)
3. In a separate PCR tube add:
    a. 18µL NFH2O
    b. 3µL T4 PNK buffer
    c. 7µL 100uM secondary primer (TEM-1 2nd-50nm-DS at 100µM)
    d. 1µL 10mM ATP
    e. 1µL T4 PNK (10U/µL)
4. Incubate at 37°C for 1 hour. Wait to do the ssDNA template strand before doing the incubation

5. Store the phosphorylated oligos at -20°C. The day of mutagenesis, dilute phosphorylated oligos 1:4 (4µL + 12µL primer), and 2nd primer 1:20 in NFH2O (1µL primer + 19µL NFH2O)

2. ssDNA Template Strand Preparation

Prepare 5 tubes for 5 tiles.

Add the following into PCR tube:

- 0.76pmol Plasmid dsDNA -> **need 2.013µg (strain 127)**
- 2µL 10X CutSmart Buffer
- 1µL 1:10 diluted ExoIII (final concentration of 10U/µL), always dilute in 1X CutSmart Buffer (NEB)
- 1µL Nt.BbvCI (10U/µL)
- 1µL ExoI (20U/µL)
- NFH2O to 20µL final volume

PCR Program (Mut/SSTSP):

- 37°C for 1 hour

- 80°C for 20min

- 5°C hold

3. Comprehensive Codon Mutagenesis Strand 1

Add the following into each tube (100µL final volume), to the PCR tube containing the ssDNA.

- 26.7µL NFH20
- 20µL 5X Phusion HF Buffer
- 4.3µL 1:100 diluted phosphorylated mutagenic oligos
- 20µL 50mM DTT
- 1µL 50mM NAD+
- 2µL 10mM dNTPs
- 1µL Phusion High Fidelity Polymerase (2U/µL)
- 5µL Taq DNA Ligase (40U/µL)

PCR program (Mut/CCMS1) for a total of 2h45min

1. 98°C for 2min
2. 98°C for 30sec
3. 55°C for 45sec
4. 72°C for 7min

- Repeat steps 2 to 4 for 15 cycles, and add additional 4.3uL of diluted oligos at beginning of cycles 6 (40min) and 11 (80min). The cycle is written on the machine.

5. 45°C for 20min
6. 5°C hold

4. Column purification using a zymo clean and concentrate kit

Step 0: Transfer the PCR content (100µL) to a clean 1.5mL Eppendorf tube.

1. Add 5 volumes (500µL) of DNA binding buffer to each reaction and mix
2. Transfer to a zymo-spin column in a collection tube
3. Centrifuge at maximum speed (13,000) for 30 seconds and discard flow through
4. Add 200µL of DNA wash buffer to the column (make sure not to mix around the DNA binding buffer and the DNA wash buffer)
5. Centrifuge at max speed (13K) for 30 seconds and discard flow through.
6. Repeat step 4 and 5.
7. **Add 17µL NFH_2O directly to the column in a new clean 1.5mL microfuge tube and incubate at room temperature at 5 minutes**
8. Centrifuge at maximum speed (13K) for 1 minute in the centrifuge

Extra 2.5uL of purified DNA to check the concentration to calculate the yield after the first purification.

Can stop here and leave the purified DNA in the -20C freezer.

5. Degrade Template Strand

Then transfer 14µL of purified DNA product to a PCR tube, then add (20µL final volume)

- 2µL 10X CutSmart buffer

- 2µL 1:50 diluted ExoIII
- 1µL 1:10 Nb.BbvCI
- 1µL ExoI

PCR Program (Mut/DTS) for 80minutes:

- 37°C for 60min

- 80°C for 20min

- 5°C hold

6) Synthesize 2nd (complimentary) mutagenic strand

To the PCR tube, add to a final volume of 100µL (use the same tube as in step 5):

- 27.7µL NFH_2O
- 20µL 5X Phusion HF buffer
- 3.3µL 1:20 diluted phosphorylated secondary primer
- 20µL mM DTT
- 1µL 50mM NAD+
- 2µL 10mM dNTPs
- 1µL Phusion HF Polymerase
- 5µL Taq DNA Ligase

PCR program (Mut/SCMS)

- 98°C for 30sec

- 55°C for 45sec

- 72°C for 10min

- 45°C for 20min

- 5°C hold

7. DNA clean up

Add into the PCR tube used in step 6:

- 2µL of DpnI (enzyme)

PCR program (Mut/DpnI) for 60minutes

8. Zymo clean and 2nd column purification.

Use the Eppendorf which has the 102 µL of plasmid oligos.

1. Add 5 volumes (500µL) of DNA binding buffer to each reaction and mix
2. Transfer to a zymo-spin column in a collection tube
3. Centrifuge at 13K speed for 30 s and discard flow-through
4. Add 200µL of DNA wash buffer
5. Centrifuge at 13K for 30 s and discard flow through
6. Repeat 4 and 5
7. **Add 8µL of NFH_2O directly to the column in a new 1.5mL microfuge tube and incubate for 5min at room temp.**
8. Centrifuge at 13K for 1min

Extra 2.5µL of purified DNA to check the concentration to calculate the yield after the second purification.

9. DNA Transformation

Transformation steps to transfer mutated DNA to XL1-blue *E.coli* cells (**Should also prepare dilution plates for sequencing to confirm your mutagenesis efficiency**).

- Use Gene Pulser Cuvette 0.1cm electrode to introduce plasmid. Both cuvettes need to be cold.
- Thaw the XL-1 blue cells in ice.
- Add 40µL of bacterial cells to the mutant(correct) to DNA sample (6µL).
- For the electroporation (knock out the bacteria to introduce the plasmids), turn on and use Ec1 for XL1 which uses 0.1cm cuvettes
- Transfer the bacteria+DNA mix to cuvettes into the middle lane (slot). Tap to equilibrate, and dry the bottom and metal sections before using the electroporation machine.
- Prepare SOC medium (960µL)

- Insert the cuvette by the protruding slot in the machine. Press pulse and hold until beep. Add SOC directly and resuspend by mixing in the corners of the cuvette
- Transfer to a tube to grow
- Incubate at 225rpm and 37°C for 1 hour

Plating XL1-blue + DNA plasmid (do 4 big plates for each tile) -> Tet + Chl

- After recovery, bring the final volume of the transformation to 2mL with additional sterile media (add 1000µL of SOC medium).
- Spread on the prepared large BioAssay dish. Serial dilution can be prepared to calculate transformation efficiencies.
- Incubate overnight at 37°C.

Recovering samples from O/N culture plate

- The next day, scrape the plate using 5mL of LB.
- Vortex the cell suspension
- Extract the library plasmid dsDNA using a mini-prep kit (Qiagen recommended) of a 1mL aliquot of the cell suspension.
- Additional mini-preps can be done if large amounts of library DNA are required.

Appendix 3: NGS Sample Preparation for Tile 1

PCR Combination Protocol for the preparation of NGS samples from Kowalsky et al., 2012. The protocol consists of adding all Illumina sequencing primers and barcoding index to the samples in a single step PCR reaction.

Preparation for 1 sample (50µL):

1. 10µL of 5X HF Phusion Buffer
2. 1µL of dNTPs (10mM)
3. 2.5µL (5uM) inner FWD primer (F1 primer T1 fwd complete)
4. 2.5µL (5uM) inner REV primer (F1 primer T1 rev)
5. 2.5µL (10uM) outer FWD primer (RPI_F)
6. 2.5µL (10uM) outer REV primer (RPI_X)
7. 0.5µL Phusion HF DNA Pol
8. 10ng template plasmid
9. NFH2O to 50µL (27.5µL)

PCR combination program

98ºC	30 seconds	1 time
98ºC 55ºC 72ºC	5 seconds 15 seconds 15 seconds	Cycle: 25 times
72ºC	10 minutes	1 time
10ºC	hold	

Two primers are designed for the NGS sample preparation: a F1 primer which anneals to the 5’ end of the sample and adds the Illumina Universal sequencing index to the front of the sample; and a R1 primer which anneals to the 3’ end of the sample and adds the barcoding index used to identify the samples when demultiplexing sequences. Following the PCR, samples are purified by gel extraction.

Primer name	Sequence	Tm (°C)	G-C%	Length (bp)
F1 primer tile 1 fwd	ACACTCTTTCCCTACACGACGCTCTTCCGATCT-TTTCATTGCTAAAAC	68.3	43.8	48
R1 primer tile 1 rev	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-AAGTCATTCTGAGAATAGTGTAT	68.5	43.6	55

Table VII.Primers used to prepare the Tile 1 samples for NGS.