In the Driver's Seat: Counterdiabatic Driving to Optimize Bacterial Infection and Cancer Treatments

An honors thesis presented for the degree of Bachelor's of Science

Jacob Marglous
Department of Physics
Brown University

Thesis Advisor: Professor Daniel Weinreich
Concentration Advisor: Professor James Valles

Abstract

Evolution of drug resistance poses a challenge to the successful treatment of infectious diseases and cancer because variants of bacterial pathogens and cancers that are resistant to front-line treatments tend to emerge rapidly. In response, clinicians can develop multistep evolutionarily-informed treatments that promote drug susceptibility. However, the efficacy of such treatments has been limited by the stochasticity of evolution and lack of *a priori* knowledge of the possible evolutionary paths a population may take. Recent work has established that increased control over an evolving population can be achieved via counterdiabatic driving, a physical technique originally developed to maintain quantum systems in a ground state as their environment is manipulated in finite time. In the evolutionary context, this has translated to maintaining a population as close as possible to the equilibrium distribution of genotypes by manipulating applied drug dosages. Here, I build on this previous work by adding temperature as a second parameter for counterdiabatic driving protocols. Ultimately, this work will hopefully lead to more efficient and successful treatments against evolving diseases.

Contents

1 Introduction

The power of evolution is evident in the vast diversity of life on Earth. The "blind watchmaker" is responsible for the emergence of living forms over millions of years from single-celled prokaryotes to the vast and constantly changing variety recognizable now. Tendency toward diversity has been called biology's "first law" [28], and evolution by natural selection, which acts upon that diversity, "perhaps biology's only guiding principle" [3]. British naturalist Charles Darwin famously stated his law of natural selection in his 1859 manuscript "On the Origin of Species" [10]. Fundamentally, the theory of natural selection states that traits emerging randomly confer fitness effects to individuals in a population. Individuals with higher fitness can produce more offspring, resulting in an increase in frequency of fitness-conferring traits over the entire population.

This power means that evolution can also pose a danger to human health on shorter time scales. The well-reported and growing problem of antibiotic resistance is fundamentally a problem of evolution. When a course of antibiotics is not prescribed correctly or not completed as prescribed, individual pathogens that happen to be resistant to the treatment by virtue of some mutation or other natural variation gain a relative fitness benefit from the treatment as they can continue to reproduce. Eventually this results in the entire population of bacteria developing resistance to the treatment. Pesticide resistance is a significant problem in agriculture, with costs of xx per year [source]. Increasing prescription of antibiotics have also raised concerns about bacterial antibiotic resistance in human pathogens. A majority of World Health Organization member regions have reported at least 50% resistence in common bacterial strains like *Escherichia coli*, *Staphylococcus aureus*, and *Streptococcus pneumoniae* [40], while a clinical class of "reserve" antibiotics has been set aside for use only in last-resort circumstances, to preserve their effectiveness in the face of growing

resistance [39].

Cancer, too, is fundamentally a disease of evolution. Tumours are highly heterogenous, and the cancer genome is extremely unstable [19] [30]. As a result, cancers generate a large number of variants that serve as fodder for natural selection. A large majority of cancer deaths result from metastasis of tumors from their original location to new tissues, which occurs when the cancer acquires, via mutation, biomarkers that allow it to seed in those new locations.[14] This high mutation rate makes treating cancer very challenging, as treatments often achieve remission but are unsuccessful at killing all the cancer cells in the body. The remaining treatment-resistant cells proliferate, resulting in recurrences of the disease. [23]

In a more emergent example, the recent COVID-19 epidemic has brought evolutionary questions into the forefront of public interest. High global case counts of coronavirus disease provide many replication cycles for the virus to generate mutations [7]. Once fitter variants emerge in the population, they gain a selection advantage over the remainder of the viral population by virtue of their enhanced transmissibility and quickly become dominant, and could even potentially evade currently available vaccines. [15]

An array of evolutionarily-informed treatments have emerged to improve treatment outcomes for such evolving pathogens. Populations will always tend towards resistance against a treatment, but clinicians can turn this tendency to their advantage by applying a well-chosen sequence of treatments to the newly-resistent population. That is, clinicians may hope that developing resistance to one drug may induce susceptibility in a second (or third, etc.). For example, to maximize treatment efficacy against cancer, clinicians have developed dosage plans utilizing combinations of multiple drugs [34].

Evolutionarily-inspired treatments have been explored experimentally *in vitro*

by testing bacterial cultures for resistance to pairs of sequentially applied antibiotics. [32]. It was found that a large majority of drug pairs promote cross-resistance, rather than drug susceptibility, so naively chosen multi-drug combinations are unlikely to result in successful treatment in the long-term.

The first reason for this is biochemical: there are a limited number of aspects of bacterial cell biology that are not conserved in humans and therefore make good potential antibiotic targets. For example, many drugs inhibit growth by targeting the bacterial cell wall [8]. If a bacterial population has developed a mechanism of resistance to one cell wall-targeting antibiotic, it is likely that this alternative mechanism will not be disrupted by a second cell wall-targeting antibiotic either.

The second reason is the stochasticity of evolution: because random mutations provide the diversity that natural selection acts upon, evolution is fundamentally a random process. Experimentally, it has been observed that even among susceptibility-promoting pairs of drugs, the desired effect of sequential dosing is not guaranteed.[31]. This is because the population can evolve along a number of genetic pathways which are chosen from essentially at random. One could imagine that a first drug acts upon some aspect of a pathogen's biology by inhibiting a particular essential protein. There is some set of mutations in the gene encoding that protein that confers resistance to the drug. Any of those mutations could be present in a portion of the population being treated, and would equivalently become common throughout the entire population via natural selection as the treatment continues. However, only a subset of the mutations conferring resistance to drug 1 also confer resistance to drug 2, so cross-susceptibility and cross-resistance are not deterministic relationships in general.

By attempting to control bacterial [20] [42] [25] and cancer cell [4] [1] popula-

tions via evolutionary trade-offs, these works have attempted to expand extant knowledge of possible evolutionary pathways toward control of the patheways themselves. However, the stochasticity of evolution has continued to limit their efficacy. Recent efforts have attempted to extend this evolutionary control of a population to *every* point throughout a treatment using counterdiabatic (CD) driving [21]. CD driving is a technique to maintain a quantum system at its ground state throughout a change in its environment applied over finite time, but can also be applied to classical stochastic systems. Applied to models of biological evolutionary systems, it can also be used to derive optimal treatment protocols that maintain the population at equilibrium genotype distributions throughout treatment. Iram et al. used as proof-of-concept a population of 16 yeast genotypes evolving under application of a single drug. In this work, I seek to extend their results by working towards CD driving protocols to maintain at equilibrium a more complex system of 32 bacterial genotypes evolving under two control parameters: antibiotic concentration and temperature.

This thesis will consist of three main parts. First, I will present a basic overview of mathematical models of biological evolution, and of counterdiabatic driving. Then, I will detail how we developed a mathematical model of the bacterial population in question from experimental data. Finally, I will present results from numerical and simulation-based approaches to understand the evolutionary dynamics of this system. While we have not yet been able to derive a counterdiabatic driving protocol for this system, we hope to continue this work in the future, and also hope that more general counterdiabatic driving protocols will pave the way for more directly clinically relevant improvements to this technique in the future.

1.1 Fitness Landscapes and the Wright-Fisher Model

A mathematical formalism for the ideas sketched above begins with the concept of a fitness landscape, first introduced by Sewell Wright in 1932 [41]. The fitness landscape exists in a space of all available genotypes (i.e. all the possible variants in the population), such that genotypes more easily accessible to each other via mutation are closer to each other. For example, two variants separated from each other by one mutation should be adjacent, while variants separated by several mutations are further away. Then, fitness (or growthrate) of each of the variants in a given environment is represented by height of a fitness function at each point in the sequence space. The result is a rugged function with hills at high-fitness alleles, separated by valleys of low fitness. A population can be represented by a single point in the sequence space by taking a weighted average of the genotypes in the population. As fitter mutants inevitably emerge and become more common due to random variation and natural selection upon them, populations tend to climb up hills on fitness landscapes.

A population's movement on a fitness landscape can be represented by a Wright-Fisher model, first described by Wright and R.A. Fisher in 1931 [2]. In the simplest version of the Wright-Fisher (WF) model that ignores selection, reproduction is simulated by randomly selecting (with replacement the members of the population at generation $t+1$ from the members of the population at generation t. That is, for the two-genotype case (individuals are denoted by a genotype that is either i or j),

$$P_{ij} = \binom{N}{j} (\frac{i}{N})^j (1 - \frac{i}{N})^{N-j} \tag{1}$$

Where the number of individuals at generation t is N, i is the number of allele A in generation t, j is the number of allele A in generation $t+1$, and P_{ij} is the probability of the count of allele A changing from i to j. Because the

transition probability P_{ij} for generation t does not depend on values of j for any generation besides the one immediately before it, Wright-Fisher evolution is a Markov process. It can be shown that the expected value of j at each time point is simply the initial number of A in the population, and that over many time steps the variance around this expectation will reduce to 0. Also note that this model does not allow for population growth: for any $j(t)$, $i(t) = N - j$. [26]

The model can be generalized to include selection by biasing the selection to favor fitter alleles. Additionally, when the population includes more than two alleles $\{A_1, A_2, \ldots A_m\}$, the frequencies of alleles in the population is represented by a vector,

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix}$$

And P_{mn} becomes a Markov transition matrix where off-diagonal entries P_{ij} are the probability of offspring transitioning from parental allele m to offspring allele n, and the diagonal entries P_{ii} are the probabilities of offspring keeping the same allele as their parents.

1.2 The Langevin equation and the Fokker-Planck approximation in one dimension

When population size is sufficiently large and mutation rate sufficiently small, changes in gene frequencies across generations can be approximated as continuous, and the Wright-Fisher model becomes a continuous Markov process. The Langevin equation provides an update formula for continuous random variables. In one dimension, the standard form Langevin equation is (2):

$$x(t+dt) = x(t) + A(x(t),t)dt + D^{1/2}(x(t),t)\mathcal{N}(t)(dt)^{1/2} \tag{2}$$

Where X is a continuous Markov random variable, $A(X,t)$ and $D(X,t)$ are smooth functions, $\mathcal{N}(t)$ is a unit normal random variable with the property that $\mathcal{N}(t)$ is independent of $\mathcal{N}(t')$ for $t \neq t'$. This is a generalization of the familiar update formula for X from deterministic one-dimensional calculus that

$$x(t+dt) = X(t) + A(x(t),t)dt \tag{3}$$

Where

$$A(x(t),dt) = \frac{dx}{dt}|_t$$

That is, the additional term introduces indeterminacy to the update formula for X via the normal distribution. [16] Paul Langevin initially used his equation to describe the velocity of diffusing particles. Newton's second law for such a particle states that

$$m\frac{dV(t)}{dt} = -\gamma V(t) + F(t)$$

Where the right-hand side of the equation is the sum of forces experienced by the particle. $-\gamma V(t)$ is a drag force with γ the drag coefficient, and $F(t)$ is a randomly fluctuating force representing the tendency of diffusing particles to spread via Brownian motion (i.e. via a continuous random walk). Comparing this specific case with the general form of the Langevin equation, $A(x,t)$ be named the drift function, representing a global shift in the position of the population of particles, and $D(x,t)$ as the diffusion function, representing the population's increasing dispersal with time.

The forward Fokker-Planck equation describes the density function of the

random variable x:

$$\frac{\partial}{\partial t}P(x;t) = -\frac{\partial}{\partial x}[A(x,t)P(x;t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[D(x,t)P(x;t)] \tag{4}$$

And the reverse Fokker-Planck equation states that:

$$-\frac{\partial}{\partial t_0}P(x,t|x_0,t_0) = A(x_0,t_0)\frac{\partial}{\partial x_0}P(x,t|x_0,t_0) + \frac{1}{2}D(x_0,t_0)\frac{\partial^2}{\partial x_0^2}P(x,t|x_0,t_0) \tag{5}$$

While these equations appear complicated, a helpful derivation is given in [36].

In a foundational paper published in 1955, Motoo Kimura derived expressions for $A(x,t)$ and $D(x,t)$ for the Wright-Fisher model in one dimension (i.e. with two competing genotypes, so one degree of freedom.) In his formulation, the drift function is:

$$A(x;t) = sx(1-x)$$

Where $x(t)$ is the frequency of allele 1, and s is a selection coefficient, or rescaled fitness value. More specifically, in this case it is half the difference in fitness between allele 1 and allele 2. For the diffusion function,

$$D(x;t) = \frac{x(1-x)}{2N}$$

Where N is the number of individuals in the population. So, the forward Fokker-Planck equation becomes [24]:

$$\frac{\partial}{\partial t}P(x;t) = -s\frac{\partial}{\partial x}x(1-x)P + \frac{1}{4N}\frac{\partial^2}{\partial x^2}x(1-x)P$$

1.3 The Langevin equation and the Fokker-Plank approximation in multiple dimensions

Now, instead of a single random variable x, consider a continuous Markov process where the state of the system at each time point is represented by a vector of M random variables

$$\vec{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \cdots \\ x_M(t) \end{pmatrix}$$

The Langevin equation becomes [17]

$$x_i(t+dt) = x_i(t) + A_i(\vec{x}(t),t)dt + \sum_{j=1}^{M} b_{ij}(\vec{x}(t),t)\mathcal{N}_j(t)(dt)^{\frac{1}{2}} \tag{6}$$

This updating formula requires a set of M A functions and M^2 b functions. Again, the element of stochasticity enters only in the final term of the equation. If each of the M^2 functions b_{ij} go to 0, then the result is a more familiar set of coupled first-order ordinary differential equations, so this equation can be interpreted as a stochastic generalization of the more familiar deterministic result. In this case, the probability density function of the state of the system as it evolves is now represented by multivariate Fokker-Planck equation. Following the derivation in [17], the multivariate forward Fokker-Planck equation is:

$$\begin{aligned} \frac{\partial P(\vec{x},t|\vec{x}_0,t_0)}{\partial t} = &-\sum_{i=1}^{M} \frac{\partial}{\partial x_i}[A_i(\vec{x},t)P(\vec{x},t|\vec{x_0},t_0)] + \\ &\frac{1}{2}\sum_{i=1}^{M} \frac{\partial^2}{\partial x_i^2}[D_i(\vec{x},t)P(\vec{x},t|\vec{x_0},t_0)] \\ &+ \sum_{i,j=1[i<j]}^{M} \frac{\partial^2}{\partial x_i \partial x_j}[C_{ij}(\vec{x},t)P(\vec{x},t|\vec{x}_0,t_0)] \end{aligned} \tag{7}$$

The multivariate backwards Fokker-Planck equation is:

$$
\begin{aligned}
-\frac{\partial P(\vec{x},t|\vec{x}_0,t_0)}{\partial t_0} &= \sum_{i=1}^{M} A_i(\vec{x}_0,t_0)\frac{\partial}{\partial x_{0i}}[P(\vec{x},t|\vec{x}_0,t_0)] \\
&+\frac{1}{2}\sum_{i=1}^{M} D_i(\vec{x}_0,t_0)\frac{\partial^2}{\partial x_{0i}^2}[P(\vec{x},t|\vec{x}_0,t_0)] \\
&+ \sum_{i,j=1[i<j]}^{M} C_{ij}(\vec{x}_0,t_0)\frac{\partial^2}{\partial x_{0i}\partial x_{0j}}[P(\vec{x},t|\vec{x}_0,t_0)]
\end{aligned} \tag{8}
$$

At each time step, each element of the vector $\vec{x}$ is changed by an increment

$$\delta x_i(dt;\vec{x},t) = x_i(t+dt) - x_i(t). \tag{9}$$

Where δx_i is a random variable. The M A_i values are means of δx_i for each i:

$$\langle \delta x_i(dt;\vec{x},t)\rangle = A_i(\vec{x},t)dt, i = 1,...,M \tag{10}$$

The M D_i values are the variances of δx_i for each i:

$$var[\delta x_i(dt;\vec{x},t)] = (\langle \delta x_i^2\rangle - \langle \delta x_i\rangle^2) = D_i(\vec{x,t})dt; i = 1,...,M \tag{11}$$

And the $\frac{1}{2}M(M+1)$ C_{ij} values are covariances of δx_i and δx_j, where i and j index different elements of the vector.

$$cov[\delta x_i(dt;\vec{x},t), \delta x_j(dt;\vec{x},t)] = \frac{1}{2}\langle \delta x_i \delta x_j\rangle - \langle \delta x_i\rangle\langle \delta x_j\rangle = C_{ij}(\vec{x},t)dt; i < j = 1,...,M \tag{12}$$

1.4 The Wright-Fisher model in multiple dimensions

When the population includes more than two alleles $\{A_1, A_2, \dots A_M\}$, the frequencies of alleles in the population is represented by a vector,

$$\vec{x}(t) = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_{M-1} \end{pmatrix}$$

The WF model is generalized to this case by expanding P_{ij}, which becomes a Markov transition matrix where off-diagonal entries P_{ij} are the probability of offspring transitioning from parental allele i to offspring allele j, and the diagonal entries P_{ii} are the probabilities of offspring keeping the same allele as their parents. The 2 genotype model had one degree of freedom and was modelled by a one-dimensional Langevin equation; here, the m genotype model has $m-1$ degrees of freedom, so it is modelled by an $m-1$-dimensional Langevin equation. (The frequency i_M can always be trivially obtained by computing $1 - \sum_{i=1}^{M-1} x_i$.)

We can simplify the multivariate Fokker-Planck equations by defining a diffusivity matrix that includes both variances of each dx_i and covariances of dx_i and dx_j. $D_{ij}(\vec{x})$ is an $M-1 \times M-1$ matrix with entries

$$D_{ij} = \frac{\langle \delta x_i \delta x_j \rangle - \langle \delta x_i \rangle \langle \delta x_j \rangle}{2}$$

Then, the forward Fokker-Planck equation (7) becomes

$$\frac{\partial}{\partial t} p(\vec{x}, t) = -\sum_{i=1}^{M-1} \frac{\partial}{\partial i} v_i(\vec{x}; \lambda(t)) p(\vec{x}, t) + \sum_{i=1}^{M-1} \sum_{j=1}^{M-1} \frac{\partial}{\partial i} \frac{\partial}{\partial j} D_{ij} p(\vec{x}, t)) \qquad (13)$$

Note that here we have replaced A_i with v_i to denote that this term represents the *velocity* of the population as it evolves toward fitter genotypes. This first term also depends on experimental parameters λ; for example, these could be temperature or drug concentration. The experimental parameters can be controlled and can vary with time; thus, they are also called *control parameters*. The first term of the Fokker-Planck equation includes this λ-dependence because the average direction of δx_i over the entire ensemble of possibilities is determined by its fitness at the given control parameters. The second term does not include a λ-dependence because it describes the stochasticity of evolution. That is, fitness-increasing and fitness-decreasing mutations are equally likely to arise, so on short time scales an evolutionary process is essentially random.

More specifically,[5]

$$v_i(\vec{x};\lambda(t)) = \sum_{j=1,j\neq i}^{M-1} m_{ij}x_j + g_{ij}(\vec{x})s_j(\lambda(t)) \tag{14}$$

and

$$D_{ij}(\vec{x},t) = \frac{g_{ij(\vec{x})}}{2N} \tag{15}$$

Where m is an $M \times M$ mutation rate matrix with off-diagonal entries m_{ij} equal to the probability of an individual mutating from i to j, and diagonal entries m_{ii} equal to negative of the total probability of mutating away from i (see Appendix 2). The matrix g has off-diagonal elements $g_{ij} = -x_i x_j$ and diagonal elements $g_{ii} = -x_i(1-x_i)$. s_j is an entry in a vector of selection coefficients of length $M-1$; as in the original description by Kimura, the selection coefficients are simply rescaled growthrates, so each genotype's selection coefficient corresponds to its fitness and its contribution toward the mean velocity of the overall population distribution.

This equation can also written in a simpler form by combining v and D into

a single Fokker-Planck operator $\mathcal{L}(\lambda(t))$:

$$\frac{\partial}{\partial t}p(\vec{x},t) = \mathcal{L}(\lambda(t))p(\vec{x},t) \tag{16}$$

From this form, it is simple to see that there is an equilibrium probability distribution associated with the operator $\mathcal{L}$, and therefore with the control parameters $\lambda(t)$. This equilibrium distribution is denoted as ρ.

$$0 = \mathcal{L}(\lambda(t))\rho(\vec{x},t) \tag{17}$$

The equilibrium probability distribution of the genotype vector can be approximated as a multivariate normal distribution.With this approximation and a continuity argument governing how the equilibrium probability distribution can change in successive time steps (see [21]), the following equation including first moments of the distribution can be defined:

$$0 \approx \sum_{\mu=1}^{M-1} m_{i\mu}\Sigma_{j\mu}(\lambda(t)) + \sum_{k=1}^{M-1}[\bar{x}_i(\lambda(t)\delta_{ik} - \bar{x}_i(\lambda(t))\bar{x}_k(\lambda(t))]s_k(\lambda(t)) \tag{18}$$

Where δ_{ik} is the Kronecker delta. This equation can then be solved for all $i, i = 1, ...M-1$ to obtain the equilibrium distribution for any landscape.

1.5 Fitness Seascapes and treatment paths in parameter space

I would now like to explore more deeply two ideas which I previously presented: the fitness landscape and the dependence of the Langevin and Fokker-Planck equations on experimental control parameters, denoted by λ. The connection between the two is simple: For some values of control parameters (i.e. drug con-

centration or temperature), each genotype in a population will have some growth rate, or fitness. These fitnesses are what give shape to a fitness landscape, and also what determine the location of the equilibrium genotype distribution in genotype space.

However, over the course of a clinical treatment of a bacterial infection, the control parameters are not static: one can imagine that the concentration of drug seen by bacteria in the bloodstream might begin at 0, before the infection is treated. Then, it might increase steadily as the patient starts a prescribed course of antibiotics before leveling off at a high steady-state concentration. (One could also imagine that it might fluctuate, dipping between daily doses or if a patient skips a dose). Because the control parameters change continuously as a function of time, the landscape the population evolves on and the equilibrium distribution of the population do as well. The result is that a static landscape of fitness hills and valleys at every time point becomes a dynamic seascape of moving fitness crests and troughs.[29]

If a population is allowed to evolve on a single landscape for infinite time, it will eventually reach the equilibrium distribution. The time required to reach equilibrium can be long because populations evolve on the time scale of generations: it takes some time for weaker mutants to die out and allow fitter ones to emerge as more frequent. Additionally, fitness landscapes can be complex, and mutations, both fitter and less fit, emerge randomly, so it can take many generations for populations to find peaks of fitness landscapes.[22]

In the case of a treatment protocol and a population evolving on a fitness seascape, the situation is even more dire. Not only is there a delay associated with the evolution of the population on the landscape at each time step of the protocol, but the landscape itself changes, resulting in an even larger lag.

1.6 Counterdiabatic driving of a Wright-Fisher population

A quantum system can exist only in discrete eigenstates which are determined by the conditions of the system. For example, a particle in a box, whose wavefunction is represented by Schrödinger's equation, is constrained to specific energy states determined by the width of the well. If the width of the well is changed, the wavefunction initially in the ground state will decompose into a superposition of higher-energy wavefunctions. The quantum adiabatic theorem states that to maintain a quantum system at the ground state throughout such a change in environmental conditions, the change must be applied infinitely slowly.

Counterdiabatic (CD) driving is a technique developed to evolve a quantum system at ground state without requiring infinite time.[12][13][18]Though successful for quantum systems [6][9], it is also possible to apply this technique to stochastic classical systems due to similaritities between the Fokker-Planck equations which govern such systems, and the Schrodinger equation.[27].

In the case of an evolutionary system, sketching the desired behavior of a CD driving protocol is simple. Rather than considering a FP operator that determines the behavior of an *actual* probability distribution in time (13), we consider a modified FP operator $\tilde{\mathcal{L}}$, dependent on the control parameters and their rate of change, which determines the behavior of the equilibrium distribution:

$$\frac{\partial}{\partial t}\rho(\vec{x};\lambda(t)) = \tilde{\mathcal{L}}(\lambda(t),\dot{\lambda}(t))\rho(\vec{x};\lambda(t)) \tag{19}$$

For a given "naive" drug treatment, determining the form of $\tilde{\mathcal{L}}$ is a two-step process. Here and elsewhere in this thesis I rely heavily on the work of Iram et al., to whom I am greatly indebted. First, a change in the vector of selection coefficients $\delta\vec{s}(t)$ necessary to maintain the population as close to equilibrium

as possible is calculated. Then, the perturbation in the original treatment plan necessary to produce those selection coefficients is determined.

Note that in (13) the variance term is not dependent on the control parameters; the CD protocol can only be implemented by modifying the selection coefficients by changing the control parameters, so the variance term is the same in $\tilde{\mathcal{L}}$ as in $\mathcal{L}$. The difference between the two operators is in the modified velocity terms $\tilde{v_i}$:

$$\tilde{v_i}(\vec{x};\lambda(t),\dot{\lambda}(t)) = v_i(\vec{x};\lambda(t)) + g_{ij}(\vec{x})\delta\tilde{s}_j(\vec{x};\lambda(t),\dot{\lambda}(t) \tag{20}$$

The whole modified Fokker-Planck equation is:

$$\frac{\partial}{\partial t}\rho(\vec{x};\lambda(t)) = -\sum_{i=1}^{M-1}\frac{\partial}{\partial i}\tilde{v}_i\rho + \sum_{i=1}^{M-1}\sum_{j=1}^{M-1} D_{ij}\rho \tag{21}$$

Where I've left out the dependencies on $\vec{x}, \lambda$ and $\dot{\lambda}$ for clarity. This can be rewritten as

$$\frac{\partial}{\partial t}\rho(\vec{x};\lambda(t)) = \mathcal{L}\rho - \sum_{i=1}^{M-1}\rho g_{ij}\delta\tilde{s}_j \tag{22}$$

And ρ is the equilibrium distribution of $\mathcal{L}$, by definition $\mathcal{L}\rho = \frac{\partial}{\partial t}\rho = 0$, so

$$\frac{\partial}{\partial t}\rho(\vec{x};\lambda(t)) = -\sum_{i=1}^{M-1}\rho(\vec{x};\lambda(t))g_{ij}(\vec{x})\delta\tilde{s}_j(\vec{x};\lambda(t),\dot{\lambda}(t)) \tag{23}$$

So the $\delta\vec{s}$ that satisfies this equation (and a continuity requirement) is the counterdiabatic driving protocol required, in terms of selection coefficients.

The second step is to convert this perturbation of selection coefficients to a perturbation of the path through the parameter space. This is a system-specific question which we have not yet considered for our model, but the supplementary information of [21] includes a helpful discussion as well as a couple of examples.

2 Methods and Results

2.1 Constructing a smooth bivariate fitness function

β-lactam antibiotics like penicillin are broken down by a bacterial protein named β-lactamase. Resistance to such antibiotics emerges via a series of 5 point mutations in the β-lactam-encoding gene: four in the coding region of the gene, and one upstream of the coding region, which influences gene transcription. A β-lactamase resistance landscape thus contains 2^5 variants, resulting from the 5 locations being either wild-type (WT) or mutant.[38]

My efforts on this project begin with growth rate data on the variants of *E. coli* that make up this landscape. For convenience, and following previous authors, I have adopted a five-bit notation, in which a WT state is assigned a 0 and a mutant a 1. For example, the wild-type genotype is denoted 00000, and the fully-mutant genotype 11111. There is a finite mutation rate that sets distances in sequence space: a single organism will acquire one mutation faster than it acquires two. Using the bit notation, we can diagram the distances between any pair selected from the 32 alleles in sequence space (figure 2.1)

In this drug-resistance model system, a useful measure of fitness is the minimum concentration of antibiotic required to stop a colony of a given variant of the bacteria from reproducing. This measure is called the minimum inhibitory concentration (MIC). A previous member of the Weinreich lab measured MIC for several replicates of populations of each of the 32 alleles at 6 temperatures (20, 25, 30, 35, 37, and 41$^\circ C$). (Table 4.1)

Our goal was to explore counterdiabatic driving on a fitness seascape in which temperature and drug concentration were both available as control parameters, building on previous work which used drug concentration alone. [21] To obtain fitness as a smooth function of both temperature and drug concentration, we first sought fitness functions for each of these control parameters separately,

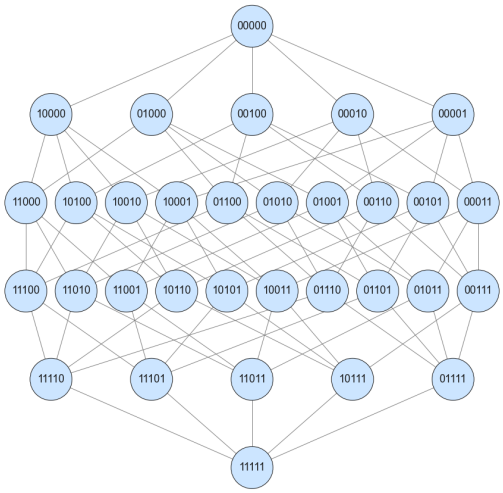

Figure 1: Bit representation of a 32-genotype landscape. This is the organization of the system in sequence space; a landscape could be visualized by plotting the fitness of each allele above its place on this map.

beginning with concentration. MIC is a limited measure of fitness because it is simply the single minimum concentration value for which there is no growth rate. Our first task was therefore to fit these MIC values to a fitness function.

Iram et al., which has served as a model for us throughout the project, used a fitness seascape created by fitting growth rates of variants of the malaria parasite with increasing drug concentration to a logistic curve. This matches matching the intuition that growth rate will be relatively unaffected by very small concentrations of drug and then quickly decline near a critical concentration. The Hill function [33]

$$g(x) = \frac{g_{drugless}}{1 + e^{\frac{IC_{50} - x}{c}}} \tag{24}$$

Takes three parameters: the drugless (maximum) growth rate $g_{drugless}$, the IC_{50}, which is the drug concentration that inhibits 50% of bacterial growth

(another fitness measure analagous to the MIC), and c, which defines the slope of the curve at the IC_{50}. Noticeably, none of these parameters are directly available to us from the *E. coli* growth data, and while the MIC represents a concentration for which there is no growth, in the Hill function g only asymptotically approaches 0. So, our first step is to argue that at the experimentally defined MIC there is most likely a small amount of growth that still occurs, and that we can therefore re-define the MIC as the IC_{95} – i.e. the drug concentration that inhibits 95% of growth. With a few lines of algebra, we can obtain a new Hill function taking IC_{95} as a parameter instead of IC_{50} (see the appendix for the full but very simple derivation).

$$g(x) = \frac{g_{drugless}}{1 + 19e^{\frac{IC_{95}-x}{c}}} \tag{25}$$

We were unable to address the other two parameters through the data as directly. First, we elected to set the $g_{drugless}$ to the same value for all of the alleles. This is somewhat of a flaw in our model. Lack of a *general* trend between fitness in the presence and absence of drug is somewhat supported by the previous fitting efforts which were the model for this work [33]. There is no easily detectable pattern between high-drug-concentration fitness and $g_{drugless}$; however, there differences in $g_{drugless}$ between alleles do exist. Further, additional research efforts in a different E. coli drug-resistance system suggests the existence of "adaptional tradeoffs" – that is, that fitness in a stressful environment is often negatively correlated with fitness in an non-stressful environment.[11]

For the final parameter c, given the lack of better options, we simply elected to fill in our model with the same value used in the previous efforts [33], $c = -.6824968$, which was constant across alleles. Because in this project we were not yet seeking to develop actionable treatment plans for E. coli but rather simply to extend counterdiabatic protocols to support a second control

parameter dimension of temperature in addition to concentration, we felt that this approximation was justified.

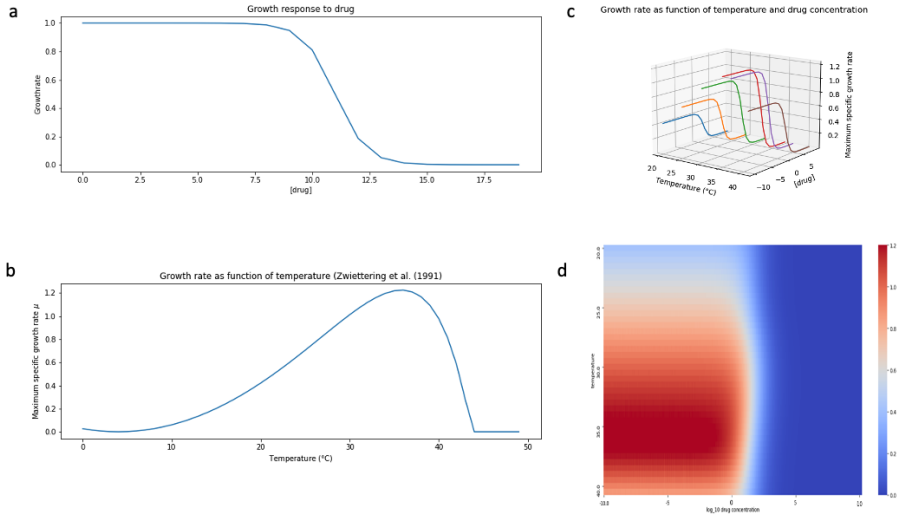

Figure 2: (a) Logistic fitness function of drug concentration, fitted to an experimental MIC. (b) Ratkowsky 3 fitness function of temperature. (c) 6 logistic fitness functions of drug concentration for allele 31 at the 6 experimental temperatures. Drugless growthrates are set by the Ratkowsky 3 model. (d) Interpolated smooth bivariate fitness function for allele 31 to fill in the missing fitness values in (c).

From an evolutionary perspective, places were fitness curves of different variants cross are particularly interesting because they are the place in parameter space where transitions between favored alleles occur. However, with the same drugless growth rate and same c, there are no such crossovers in our model along a single-temperature slice of the parameter space. An interesting future direction would be to test the drugless growth rates of each genotype to improve the realism of the model.

Turning our attention to temperature, we reviewed the available published functions of bacterial growth rate as a function of temperature. In the food sciences literature [43] we found a useful model, called the Ratkowsky 3 model,

that describes an increase in growth rate from a minimum growth-supporting temperature to the biological temperature of $37°C$, followed by a more rapid decline to a maximum growth-supporting temperature, which matches the shape one would intuitively expect for such a function. The Ratkowsky model [35]

$$g(T) = [b_3(T - T_{min})]^2 * (1 - \exp[c_3(T - T_{max})]) \tag{26}$$

takes 4 parameters which have been fitted to growth curves for *Lactobacillus plantarum*, another species of bacteria. Again, in the absence of results directly relevant to our variants of *E. coli*, we chose to fill in our model with these fitted constants as-is: $b_3 = .0410, c_3 = .161 T_{min} = 3.99, T_{max} = 43.7$. We used this function to set the maximum growth rate for the Hill function of growth rate as a function of drug concentration at each of the six experimental temperatures, for each of the 32 tested alleles. This resulted in six parallel quasi-experimental fitness-concentration curves for each of the alleles (figure 2.1). Then, to obtain a smooth fitness function of concentration and temperature $G(T, [drug])$, we applied interpolation tools from the SciPy toolkit to fill in the missing values. [37]

We defined fitness as a relative growth rate advantage over a reference allele, for which we selected allele 31 (11111). That is, for allele i, the selection coefficient s_i is

$$s_i(T, [drug]) = \frac{growthrate_i(T, [drug])}{growthrate_{31}(T, [drug])} - 1 \tag{27}$$

The subtraction simply changes the reference w_{31} from 1 to 0. With this, we successfully converted the available experimental data to a smooth fitness function of temperature and drug concentration for each allele. At low concentrations, fitnesses for each allele are very similar, but as concentrations increase,

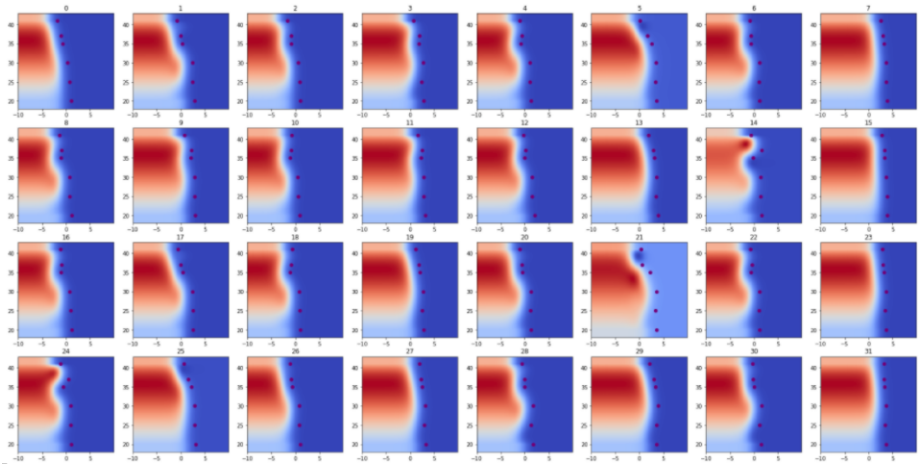

Figure 3: Smooth bivariate fitness functions of drug concentration and temperature. log[$drug$] is shown on the x axis and temperature on the y. Scale of the heatmap is the same as in (Figure 2.2). Purple dots indicate the locations of the MICs used to build the model.

interesting differences begin to emerge. Notably, for several alleles there are multiple fitness peaks and valleys in the temperature dimension, suggesting interesting opportunities for driving protocols that exploit these features to favor certain alleles over others.

2.2 Drawing treatment plans and calculating equilibrium distributions

With the construction of the fitness seascape complete, we can now consider a variety of treatment regimens in the form of pathways through the parameter space. For the moment I'll consider a simple case, in which temperature is increased as a function of time while concentration is held constant. As we make a treatment plan by drawing a path through the parameter space, we can pull out a selection coefficient at each point on the path for each allele.

We can then use those selection coefficients to calculate the equilibrium distribution associated with the control parameters at each time step. There

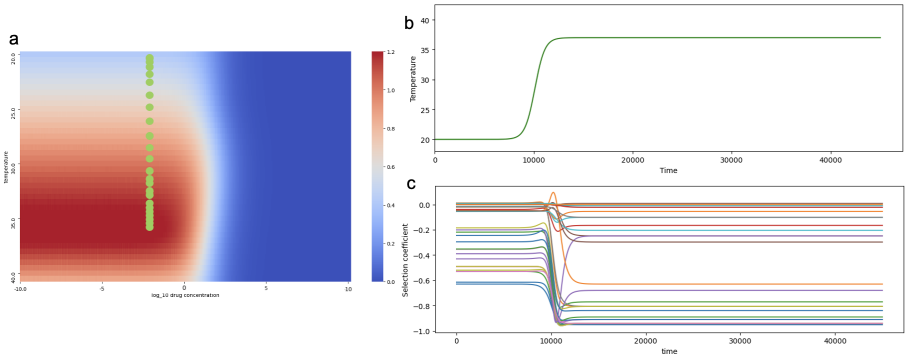

Figure 4: A path through the parameter space. (a) a temperature treatment at constant drug concentration is drawn through the parameter space as a series of green dots. (b) The temperature of the treatment is plotted as a function of time. (c) Selection coefficients of the 32 alleles are plotted at sampled points on the treatment path.

is no closed form of the equilibrium probability distribution for many alleles. Therefore, to calculate equilibrium frequencies for each time step, we solved (18) numerically. (As a reminder, this equation was derived by approximating the population as a multivariate normal distribution in sequence space and then specifying regularity conditions.) The numerical calculation was performed by a solver in Python included in the SciPy "optimize" library. After making a guess at the initial distribution manually, the solver was run at each time step, updating the guess genotype frequency vector for the solver to look near at each step with the solution for the previous step.

We were challenged by practical difficulties in the solution of this equation. (18) is underdetermined, and the solver returned unwanted negative genotype frequencies at places in treatment protocols where changes in the control parameters were most rapid. We attempted a number of approaches to produce more accurate answers, including running an agent-based simulation for long times on the landscape of the time step (see next subsection) and using the endpoint of the simulation as the guess solution for the solver. The most successful approach

was to increase the density of points along the treatment pathway in parameter space which were sampled for the equilibrium calculation. More frequent samples reduced the distance in sequence space between successive instantaneous equilibria, enabling the solver to more effectively locate the desired frequencies. Reducing the size of the genotype space also increased the effectiveness of the solver. (figure 2.2). However, this is somewhat unsatisfying, as we hope to calculate equilibria for all genotypes in the system.

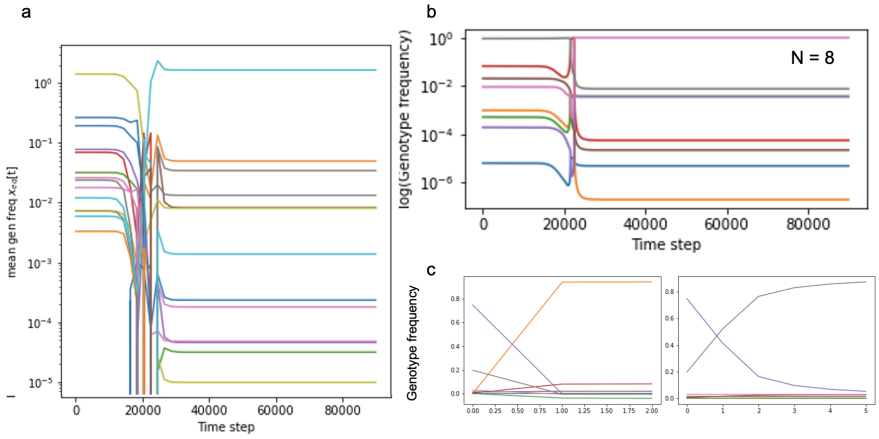

Figure 5: Troubleshooting the numerical equilibrium calculation. (a) For 32 genotypes, the solver returns negative genotype frequencies. (b) Effect of dimensionality on the solver: the solver returns more reasonable genotype frequencies for the same seascape when only 8 genotypes are considered. (c) Increasing sampling density improves the solver's accuracy: decreasing the time between samples of the control parameter path results in more reasonable answers. Left is a time difference of 1 unit; right is a time difference of .5 units.

2.3 Agent-based simulation to determine behavior of populations under treatment plans

Iram et al. used counterdiabatic driving to minimize the lag between the numerically-calculated equilibrium distribution evolving under a drug concentration ramp, and the actual frequency distribution of the population. Because

we were unable to generalize their simulation to support a 32-genotype system as well as their 16-genotype one, we used their simulation, written in C++, as a template for a new agent-based simulation in Python that can support any number of genotypes. Due to the increased computational cost of doubling the number of genotypes in the simulation, we also attempted to optimize the efficiency of the model. The model, run for M time steps, takes as inputs the initial populations of each of the N genotypes, an NxM selection matrix s containing columnwise the fitnesses of each of the alleles, an intrinsic birthrate b_0, a deathrate d, and a carrying capacity p_0. The values s_{ij} of the selection matrix are calculated by defining the drug concentration and temperature as a function of time, and evaluating the interpolated fitness function for each allele at each time step.

The simulation works as follows: at each time step, each individual has a probability of dying d, and a probability of producing a single offspring b. d is a constant and b is defined for each allele and each time step of the simulation. If the total population of the system is greater than p_0, $b = 0$. Otherwise, the birthrate is biased by the fitness of the individual's genotype: $b(i) = b_0(1 + s_i)$. Transitions between genotypes are written into the simulation via mutations in the offspring: An offspring will most likely belong to the same genotype as its parent. However, there are small and equal probabilities that it will mutate to any of the genotypes directly adjacent to it in genotype space.

The main computational cost in the original simulation were the two independently drawn random numbers to determine whether each individual died, and then whether they gave birth, at each step of the simulation. In the original simulation, a random number-generating function was called twice for each member of the population at every time step. However, we found that it was faster to generate two vectors of random numbers of length equal to the total

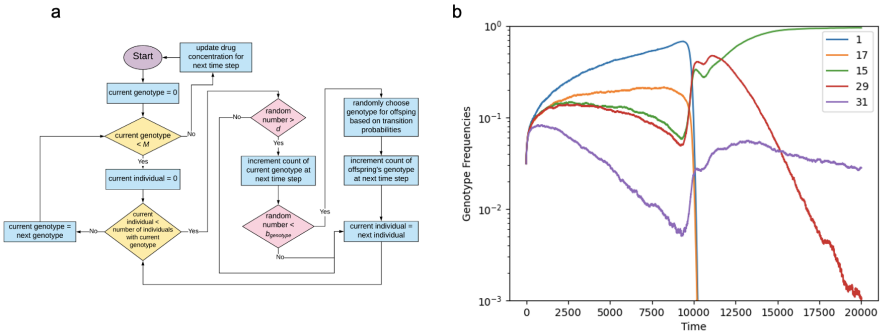

Figure 6: An agent based model of a population evolving on a seascape. (a) A flowchart of the ABM (reproduced from [21]. (b) A representative iteration of the ABM for the 32 genotype *E. coli* model on the temperature ramp in (figure 2.2). Only the trajectories of the fittest 5 alleles are plotted.

population size (one for births, and one for deaths) at each time step, and then call an element of those vectors to determine the behavior of each individual at each time step. Additionally, time-intensive loops were minimized by calculating both whether an individual will give birth, and if it gives birth, what the genotype of its offspring will be, in the same loop. Anecdotally, this optimization reduced the time required to run the simulation from about 90 minutes to about 15.

We first tested the accuracy of this simulation by using it to simulate over the paths on the 16-genotype seascape published Iram et al., and comparing its output to that of the simulation in that paper which it was based upon. As shown in (fig), it displays remarkable fidelity to the original agent-based model.

As a second test, we developed toy fitness landscapes with single clearly-fittest alleles that changed abruptly, to assess the model's ability to recognize these fittest mutants. The simulation quickly and correctly identified these mutants, with the time between successive selective sweeps increasing with the distance in sequence space between the successive favorites, as expected. We then turned our attention to the precision of the simulation for our dataset by

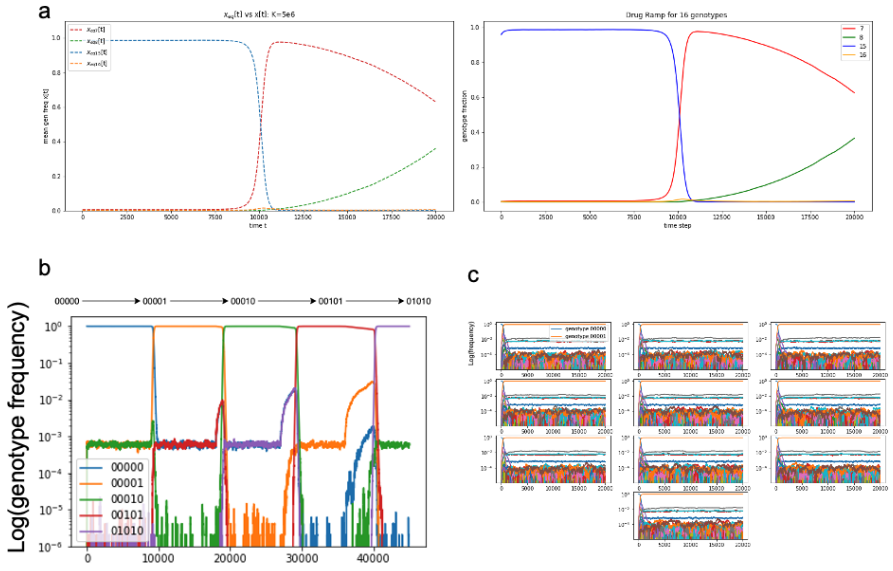

Figure 7: Validating the ABM. (a) Comparing the results of the ABM written in Python to the C++ ABM in [21]. (b) Testing the ABm on a simplified landscape where only one allele has a nonzero fitness at every timepoint. (c) 10 iterations of the ABM on the same constant landscape produce similar results.

considering the evolution of the population under constant conditions. We would expect that with sufficient time, the distribution would favor the fittest allele. In successive iterations of the simulation from the same initial genotype frequencies, the results were satisfyingly consistent. Additionally, simulations beginning from a different, and vastly distant, frequency distribution also converged to the same expected values.

Throughout this work, we consider the large population, frequent mutation regime ($\mu N > 1$). Specifically, in this simulation we consider populations on the order of $10^5 - 10^6$, and a total mutation rate (that is, the likelihood that an offspring will belong to a different genotype than its parent) of $\mu = .01$). So, at every time step a large number ($10^3 - 10^4$) of mutations will be generated. There are only 32 alleles which can be mutated into, so we expect that the

time required for the simulation to explore the sequence space sufficiently to recognize fittest alleles is low. This is supported by our controls.

We hope to follow Iram et al. and use the simulation to compare the real behavior of a population on a path through the control parameter space to its equilibrium distributions. However, these efforts are on hold as we have been unable to obtain reliable equilibrium calculations for the whole system.

3 Discussion and Conclusion

Anti-drug resistance is a challenging problem facing clinicians treating bacterial or viral infections and cancer. Resistance is driven by the high reproduction rates of these pathogenic cells, which contributes to a high degree of genetic diversity that enables these pathogenic populations to efficiently "find" fitter genotypes and evade treatment. Previous work has suggested that this problem of of evolution can be turned on its head by identifying treatment regimens that promote cross-susceptibility rather than cross-resistance. That is, taking the evolution of treatment resistance as a given, we can design multi-component treatment schedules in which early development of resistance to one portion of the treatment leads to susceptibility to another. We could even imagine a "susceptibility-inducing cycle" of several components in which resistance to A induces susceptibility to B, any surviving B resistors are susceptible to C, and any surviving C resistors are more susceptible to A.

Evolution-inspired treatments offer a number of intriguing possibilities for novel therapies. For example, the current standard of care for many cancers is to treat with the maximum tolerable dose (MTD) of available chemotherapy. However, this treatment plan frequently results in resistant tumor populations emerging in the form of recurrent disease. It also comes at great cost to the

patient, in the form of side effects from high-dose chemotherapy. Using knowledge of cross-susceptibility and the range of evolutionary responses a cancer population could make to drug exposure, treatments could be designed with the goal not to eliminate the cancer, but to maintain the population in a drug-susceptible state. In this manner, cancer could be reimagined as a chronic, but not life-threatening, disease requiring consistent, manageable, low-side-effect treatment.[23]

Understanding the in- and out-of-equilibrium dynamics of a pathogenic population is a crucial prerequisite for the implementation of an evolutionarily-motivated treatment plan. To take advantage of cross-susceptibility, one needs to be certain that the population has actually equilibrated to the first treatment before the second drug is applied. However, in general, given the stochastic nature of evolution, it could take a population a very long time to actually reach equilibrium. In fact, following the mathematical formalism described in the introduction (starting with the multivariate Langevin equation, whose probability distribution is given by a Fokker-Planck equation) we see that the adiabatic theorem states that we would have to change the control parameters infinitely slowly to keep the population at equilibrium throughout the treatment.

The advantage of counterdiabatic driving in any context (evolutionary or otherwise) is that it bypasses the adiabatic theorem, minimizing the difference between the equilibrium and actual distributions without requiring infinite equilibration. The goal of this project was to explore whether adding an additional control parameter in the form of temperature would either increase the effectiveness of driving (i.e. maintain the population even closer to equilibrium) or provide new opportunities for control – that is, to enable driving to distributions dominated by different genotypes. The second goal is particularly useful, as one can imagine that a clinician may wish to drive a pathogenic population toward

a particular susceptible allele (or alternatively, away from a particularly dangerous allele), but using only a single control parameter, there may be nowhere in the one-dimensional parameter space where that allele is favored. Increasing the dimensionality of the parameter space increases the likelihood that there is some location where a desired allele will be favored at equilibrium – and then, a counterdiabatic driving protocol will reduce the time required to reach that spot.

We have thus far been unable to actually derive or demonstrate a counterdiabatic driving protocol over the two-dimensional sequence space, in large part because we have struggled to accurately calculate equilibrium distributions to compare our simulation to. In a less rigorous sense, I think this challenge reflects a fundamental struggle at the heart of this research problem in that computational approaches to problems in evolution reflect the computational behavior of evolving populations. It could be said that populations search for maxima of fitness functions algorithmically [22]. A fitter individual has a high likelihood of producing (similar and perhaps even more fit) offspring, while a less-fit individual has a smaller but still non-zero likelihood of reproducing, in the same way a Metropolis algorithm searching for the minimum of an energy function will accept an energy-decreasing step with a high probability, and accept other steps with a small non-zero probability. In this work, I tried, and failed, to support the finicky numerical solution to the equilibrium frequencies at each time step of the simulation by simulating for a long time at the landscape of that time step, and inputting the final distribution of that simulation to the solver as an improved guess. This strategy failed because it was challenging to determine whether the population had truly equilibrated, or was simply stuck waiting for a mutationally distant but favored allele to gain a large enough population to produce a selection sweep. Running the simulation for long enough to essen-

tially guarantee that the population would truly equilibrate for each time step was practically infeasible, but if I was able to run the simulation for infinite time on the landscape of each time step of the treatment schedule, I should be able to find the equilibrium frequency. I visualize this as simulating the equilibrium frequencies along the dimension of the treatment schedule by simulating many times in a perpendicular dimension at each time step.

Developing strategies to resolve this computational complexity is a primary goal for the future of this project. It may be possible to derive reasonable equilibrium distributions for populations with more mutants using the same numerical techniques applied here, with a number of modifications. It is currently infeasible to reduce the distance between the treatment time steps for which equilibrium distributions are calculated enough for the solver to be effective. Nonethless, we hope to more fully test the parameters of the solver and achieve success on this aim. Alternatively, it may be possible to develop a "smarter" solving algorithm that detects unreasonably large changes in genotype frequencies between successive time steps and reduces the distance between successive samples of the treatment path only for such stress points, on an as-needed basis. And finally, while this approach has not worked to date, it may still be possible to deploy long-term simulations on a static landscape for time steps that are particularly challenging to characterize. If the simulations are applied only to time steps where they are necessary, they could be run for long enough to produce more useful predictions to the solver.

Alternatively, because the frequency of the genotypes in the system at each time point is updated by a Markov process, it may be possible to set aside the equilibrium distribution calculation described here and instead calculate equilibrium distributions from the stationary distribution of the Markov transition matrix at each time step. Indeed, the simulations already include as input

a transition matrix for each individual (see appendix). Without taking selection into account, because each individual behaves independently, the transition matrix for the entire population should be related to this individual transition matrix in a simple manner. It should also be possible to work out via birthrates how incorporating fitness would modify this population-level transition matrix, in a manner analogous to the way fitness biases the birthrate of each mutant in the agent-based simulation. It would be worthwhile to explore whether such a transition matrix would even be guaranteed to have a stationary distribution at each time step, and if so, whether the stationary distributions match those given by the moment-closure numerical approach.

Regardless, once the calculation of equilibrium distributions is more completely worked out we will be able to more systematically compare equilibrium distribution to simulation results and hopefully demonstrate a lag between them, in the manner of Iram et al. Then, we could potentially work toward the stated aim of this project, to demonstrate that a set of modified selection coefficients derived via counterdiabatic driving forces the simulated population distribution as close as possible to the equilibrium one.

Beyond completing the initial aims of this project, there are many additional future directions to consider. Counterdiabatic driving is far from any clinical practicality for a number of reasons. First, implementing a CD protocol requires a high degree of power over the values of the control parameters at every point in time. In a clinical context, where these parameters may fluctuate significantly due to a multitude of factors (dehydration, nutrition, disease progress, simple movement, etc.), it is hard to imagine ever attaining a useful level of control. However, an intriguing result reported by Iram et. al. was that even a suboptimal counterdiabatic driving protocol which took the same shape as the original counterdiabatic protocol but with a constraint on the maximum value

of the control parameter (suggestive of a patient's maximum tolerable dose) was effective at reducing lag from the equilibrium qenotype distribution at every time point. There is also additional ongoing research which seeks to model evolution of pathogenic populations under "real-life" pharmacodynamics including missed treatment doses and varying dosing schedules. Combining these avenues of work may enable counterdiabatic predictions that also provide useful clinical bounds on the maximum variance in the control parameters which still leads to the desired outcome.

These are challenges in real-life adherence to counterdiabatic driving prescriptions. However, there are also many opportunities for development of more accurate fitness functions themselves. To date this project has focused on fitness of a simplified bacterial population in which growth rate depends only on the control parameters. However, the ultimate goal of this project would be to optimize treatments for bacterial infections and cancer in real patients. In these cases, fitness would be a much more complicated function of position: for example, cancer cell fitnesses vary throughout the tumor. A proper counterdiabatic driving protocol in this case would have to somehow find the optimal drug concentrations at multiple locations in the body, or at several places in the bloodstream. Also, tumors often include several cell subtypes that engage in cooperation to support the growth of the tumor by promoting blood flow, nutrient acquisition, etc. A more realistic set of fitness functions could perhaps incorporate optimal treatment options to balance the fitnesses of multiple cancer cell subtypes, so that a counterdiabatic driving protocol could optimize treatment on a seascape of tumor fitnesses, rather than cell fitnesses.

Additionally, the particular focus of this work on driving across a parameter space of temperature and a single drug was driven by the available data and a desire to begin explore counterdiabatic driving in a multi-dimensional parameter

space. A useful extension of this work would be to replace temperature with a second drug, or even to add additional parameter dimensions to see if this opens up opportunities to drive even closer to the equilibrium distribution at every time step, or new opportunities to design more optimal paths in the parameter space.

Setting aside these theoretical developments, the work presented here and in Iram et. al makes predictions that can be tested experimentally. It would be interesting to apply counterdiabatic driving protocols to real bacterial populations, take frequent samples for sequencing, and observe if experimental results match the theoretical predictions for the equilibrium frequencies at each time step. I would predict that inability to exactly regulate the control parameters experimentally will result in an inability to completely replicate the equilibrium frequencies of the population throughout the entire experiment. However, even an approximation of a counterdiabatic driving protocol would likely represent an improvement over a naive protocol.

Ultimately, I hope that counterdiabatic driving will become one element of a new vanguard of evolutionarily-inspired treatments for bacterial infections and cancer. While much work remains to be done before these ideas are ready to test in the clinic, a new quantum-inspired evolutionary biology seems to be just on the horizon.

4 Appendices

4.1 Appendix 1: Measured MICs for *E. coli* mutants

Allele	20°C	25°C	30° C	35°C	37°C	41°C
00000	11.3	5.65	2	0.175	0.09	0.015635
00001	724.45	256.125	256.125	2	1	0.175
00010	11.3	8	4	0.125	0.125	0.02
00011	724.45	256.125	512.25	64.03	90.55	5.65
00100	22.65	11.3	11.3	0.175	0.25	0.09
00101	4098	2049	2897.7	362.2	45.3	1.4
00110	16.01	11.3	16.01	0.25	0.25	0.125
00111	8096	4098	4098	1448.7	1024.5	362.2
01000	16.01	5.65	5.65	0.09	0.125	0.045
01001	1024.5	724.45	724.45	128.06	181.1	5.65
01010	8	4	2.85	0.09	0.125	0.175
01011	1448.7	724.45	724.45	181.1	362.2	45.3
01100	128	32.015	32.015	1	1.4	0.5
01101	8096	2897.7	4098	1448.7	724.5	90.55
01110	45.3	22.65	22.65	0.705	45.3	0.25
01111	8096	8096	4098	2049	2049	512.25
10000	16.01	8	5.65	0.09	0.125	0.0625
10001	362.2	362.2	256.125	2.85	1	0.25
10010	11.3	8	5.65	0.09	0.125	0.25
10011	724.45	362.2	512.25	128.06	64.03	16.01
10100	22.65	16.01	22.65	0.355	0.25	0.09
10101	4098	2049	4098	181.1	4	2
10110	16.01	11.3	16.01	0.25	0.355	0.175
10111	4098	2897.7	4098	1448.7	1024.5	724.45
11000	11.3	8	8	0.25	2.85	0.0625
11001	1448.7	724.45	724.45	181.1	32.015	4
11010	11.3	5.65	5.65	0.25	0.125	0.09
11011	2049	1024.5	1448.7	362.2	181.1	90.55
11100	64.03	22.65	64.03	1	1	0.25
11101	4098	4098	8096	2049	1024.5	128.0625
11110	32.015	16.01	45.3	1	0.705	0.5
11111	4098	2897.7	4098	2049	1448.7	1448.7

Table 1: MICs for 32 *E. coli* alleles at 6 temperatures.
MICs are cefoxatime concentrations in μ g/mL. Bit notations are mutant (1) or wild-type (0) at, left to right, M4205A, A42G, E104K, M182T, G238S.

4.2 Appendix 2: Mutation rate matrix for 8 genotypes

-9E-03	3E-03	3E-03	0	3E-03	0	0	0
3E-03	-9E-03	0E+00	3E-03	0	3E-03	0	0
3E-03	0	-9E-03	3E-03	0	0	3E-03	0
0	3E-03	3E-03	-9E-03	0	0	0	3E-03
3E-03	0	0	0	-9E-03	3E-03	3E-03	0
0	3E-03	0	0	3E-03	-9E-03	0	3E-03
0	0	3E-03	0	3E-03	0	-9E-03	3E-03
0	0	0	3E-03	0	3E-03	3E-03	-9E-03

4.3 Appendix 3: Derivation of Logistic Fitness-drug concentration curve from MIC

The original MIC equation from [33] is (24):

$$g(x) = \frac{g_{drugless}}{1 + e^{\frac{IC_{50} - x}{c}}} \tag{28}$$

If the MIC is redefined $MIC = IC_{95}$, $g(MIC) = .05 g_{drugless}$:

$$.05 = \frac{1}{1 + e^{\frac{IC_{50} - IC_{95}}{c}}}$$

$$IC_{50} = c \ln 19 + IC_{95}$$

$$g(x) = \frac{1}{1 + e^{\ln 19 + \frac{IC_{95} - x}{c}}}$$

$$g(x) = \frac{1}{1 + 19 e^{\frac{IC_{95} - x}{c}}} \tag{29}$$

5 Acknowledgements

There are many people who deserve much more than the few words of thanks I can offer here, but I'll try nonetheless.

First, my academic advisors: thank you to Professor Dan Weinreich, for mentoring me for this thesis project. We may have only met in-person once, but I look forward to many more in the future. Many thanks as well to the other members of the Weinreich Lab, Maya Weissman, Dave Morgan, and Eugene Raynes, for taking a chance on me!

Thank you to Professor Jim Valles, for advising me on this project, for employing me as a TA, and for (re)introducing me to physics. I would not be here without you!

Thank you to Professor Nicolas Fawzi, for introducing me to biophysics and teaching me how to read and think like a scientist. Thanks as well to all the members of the Fawzi Lab, but especially Kaylee Mathews and Scott Watters, for putting up with my many mistakes at the lab bench.

Thank you to Professor Susan Gerbi, Dr. Miiko Sokka, and Jacob Bliss for welcoming me into their lab before I even had a Brown ID card, and to Dr. Thomas Kupper for introducing me to the world of science. Thank you to Dr. Serena Lofftus for teaching me how to pipette, and to Dr. Tian Tian for convincing me I was cut out to medical school.

Many thanks for the support of our collaborators on this project at Case Western University and the Cleveland Clinic: Dr. Jacob Scott and Professor Mike Hinczewski. In particular, thank you to Drs. Nikhil Krishnan and Shamreen Iram for their generous contributions of time, data, and code.

Thank you to the friends I've made in my time at Brown: Patrick, Oscar, Ben, Allie, Sarah, and Savannah – I can't imagine the last four years without you. And to Alex – I hope we'll finally be back in the same city soon!

Finally, to my family: my mother, Linda, my sister, Sam, of course, my father, David. You will always be my inspiration.

References

[1] Ahmet Acar et al. "Exploiting evolutionary steering to induce collateral drug sensitivity in cancer". en. In: *Nature Communications* 11.1 (Dec. 2020), p. 1923. ISSN: 2041-1723. DOI: 10.1038/s41467-020-15596-z. URL: http://www.nature.com/articles/s41467-020-15596-z (visited on 04/03/2021).

[2] Nicolas BacaÃ«r. *A Short History of Mathematical Population Dynamics*. en. London: Springer London, 2011. ISBN: 978-0-85729-114-1 978-0-85729-115-8. DOI: 10.1007/978-0-85729-115-8. URL: http://link.springer.com/10.1007/978-0-85729-115-8 (visited on 04/03/2021).

[3] Philip Ball. *The Computational Foundation of Life*. en. URL: https://www.quantamagazine.org/the-computational-foundation-of-life-20170126/ (visited on 04/02/2021).

[4] David Basanta, Robert A. Gatenby, and Alexander R. A. Anderson. "Exploiting Evolution To Treat Drug Resistance: Combination Therapy and the Double Bind". en. In: *Molecular Pharmaceutics* 9.4 (Apr. 2012), pp. 914–921. ISSN: 1543-8384, 1543-8392. DOI: 10.1021/mp200458e. URL: https://pubs.acs.org/doi/10.1021/mp200458e (visited on 04/03/2021).

[5] G.J. Baxter, R.A. Blythe, and A.J. McKane. "Exact solution of the multi-allelic diffusion model". en. In: *Mathematical Biosciences* 209.1 (Sept. 2007), pp. 124–170. ISSN: 00255564. DOI: 10.1016/j.mbs.2007.01.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S002555640700003X (visited on 03/23/2021).

[6] M V Berry. "Transitionless quantum driving". en. In: *Journal of Physics A: Mathematical and Theoretical* 42.36 (Sept. 2009), p. 365303. ISSN: 1751-8113, 1751-8121. DOI: 10.1088/1751-8113/42/36/365303. URL: https://iopscience.iop.org/article/10.1088/1751-8113/42/36/365303 (visited on 11/25/2020).

[7] Christina Burch et al. *Perspective: The evolutionary dangers of high COVID case counts*. en. preprint. EcoEvoRxiv, Feb. 2021. DOI: 10.32942/osf.io/tmjrp. URL: https://osf.io/tmjrp (visited on 04/02/2021).

[8] K. Bush. "Antimicrobial agents targeting bacterial cell walls and cell membranes: -EN- -FR- Les agents antimicrobiens ciblant la paroi cellulaire et la membrane cellulaire des bactÃ©ries -ES- Agentes antimicrobianos que atacan la pared y la membrana celulares de las bacterias". In: *Revue Scientifique et Technique de l'OIE* 31.1 (Apr. 2012), pp. 43–56. ISSN: 0253-1933. DOI: 10.20506/rst.31.1.2096. URL: https://doc.oie.int/dyn/portal/index.seam?page=alo&aloId=31367 (visited on 04/03/2021).

[9] Adolfo del Campo. "Shortcuts to Adiabaticity by Counterdiabatic Driving". en. In: *Physical Review Letters* 111.10 (Sept. 2013). ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.111.100502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.111.100502 (visited on 01/05/2021).

[10] Charles Darwin. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, 1859.

[11] Suman G Das et al. "Predictable properties of fitness landscapes induced by adaptational tradeoffs". In: *eLife* 9 (May 2020). Ed. by Richard A Neher et al. Publisher: eLife Sciences Publications, Ltd, e55155. ISSN: 2050-084X. DOI: 10.7554/eLife.55155. URL: https://doi.org/10.7554/eLife.55155 (visited on 11/17/2020).

[12] Mustafa Demirplak and Stuart A. Rice. "Adiabatic Population Transfer with Control Fields". en. In: *The Journal of Physical Chemistry A* 107.46 (Nov. 2003), pp. 9937–9945. ISSN: 1089-5639, 1520-5215. DOI: 10.1021/jp030708a. URL: https://pubs.acs.org/doi/10.1021/jp030708a (visited on 11/25/2020).

[13] Mustafa Demirplak and Stuart A. Rice. "Assisted Adiabatic Passage Revisited â ". en. In: *The Journal of Physical Chemistry B* 109.14 (Apr. 2005), pp. 6838–6844. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/jp040647w. URL: https://pubs.acs.org/doi/10.1021/jp040647w (visited on 11/25/2020).

[14] Hanna DillekÃ¥s, Michael S. Rogers, and OddbjÃžrn Straume. "Are 90% of deaths from cancer caused by metastases?" In: *Cancer Medicine* 8.12 (Aug. 2019), pp. 5574–5576. ISSN: 2045-7634. DOI: 10.1002/cam4.2474. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6745820/ (visited on 04/02/2021).

[15] Arnaud Fontanet et al. "SARS-CoV-2 variants and ending the COVID-19 pandemic". English. In: *The Lancet* 397.10278 (Mar. 2021). Publisher: Elsevier, pp. 952–954. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(21)00370-6. URL: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)00370-6/abstract (visited on 04/02/2021).

[16] Daniel T. Gillespie. "The mathematics of Brownian motion and Johnson noise". en. In: *American Journal of Physics* 64.3 (Mar. 1996), pp. 225–240. ISSN: 0002-9505, 1943-2909. DOI: 10.1119/1.18210. URL: http://aapt.scitation.org/doi/10.1119/1.18210 (visited on 12/31/2020).

[17] Daniel T. Gillespie. "The multivariate Langevin and FokkerâPlanck equations". en. In: *American Journal of Physics* 64.10 (Oct. 1996), pp. 1246–1257. ISSN: 0002-9505, 1943-2909. DOI: 10.1119/1.18387. URL: http://aapt.scitation.org/doi/10.1119/1.18387 (visited on 12/31/2020).

[18] D. GuÃ©ry-Odelin et al. "Shortcuts to adiabaticity: Concepts, methods, and applications". en. In: *Reviews of Modern Physics* 91.4 (Oct. 2019). ISSN: 0034-6861, 1539-0756. DOI: 10.1103/RevModPhys.91.045001. URL: https://link.aps.org/doi/10.1103/RevModPhys.91.045001 (visited on 01/05/2021).

[19] Caitriona Holohan et al. "Cancer drug resistance: an evolving paradigm". en. In: *Nature Reviews Cancer* 13.10 (Oct. 2013), pp. 714–726. ISSN: 1474-175X, 1474-1768. DOI: 10.1038/nrc3599. URL: http://www.nature.com/articles/nrc3599 (visited on 04/02/2021).

[20] L. Imamovic and M. O. A. Sommer. "Use of Collateral Sensitivity Networks to Design Drug Cycling Protocols That Avoid Resistance Development". en. In: *Science Translational Medicine* 5.204 (Sept. 2013), 204ra132–204ra132. ISSN: 1946-6234, 1946-6242. DOI: 10.1126/scitranslmed.3006609. URL: https://stm.sciencemag.org/lookup/doi/10.1126/scitranslmed.3006609 (visited on 04/03/2021).

[21] Shamreen Iram et al. "Controlling the speed and trajectory of evolution with counterdiabatic driving". en. In: *Nature Physics* (Aug. 2020). ISSN: 1745-2473, 1745-2481. DOI: 10.1038/s41567-020-0989-3. URL: http://www.nature.com/articles/s41567-020-0989-3 (visited on 09/21/2020).

[22] Artem Kaznatcheev. "Complexity of evolutionary equilibria in static fitness landscapes". In: *arXiv:1308.5094 [cs, q-bio]* (Aug. 2013). arXiv: 1308.5094. URL: http://arxiv.org/abs/1308.5094 (visited on 04/01/2021).

[23] Artem Kaznatcheev et al. "Cancer treatment scheduling and dynamic heterogeneity in social dilemmas of tumour acidity and vasculature". en. In: *British Journal of Cancer* 116.6 (Mar. 2017), pp. 785–792. ISSN: 0007-0920, 1532-1827. DOI: 10.1038/bjc.2017.5. URL: http://www.nature.com/articles/bjc20175 (visited on 07/06/2020).

[24] Motoo Kimura. "Kimura_StochasticProcesses.pdf". In: vol. 20. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1955, pp. 33–53. DOI: 10.1101/SQB.1955.020.01.006.

[25] Jeff Maltas and Kevin B. Wood. "Pervasive and diverse collateral sensitivity profiles inform optimal strategies to limit antibiotic resistance". en. In: *PLOS Biology* 17.10 (Oct. 2019). Ed. by Hinrich Schulenburg, e3000515. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.3000515. URL: https://dx.plos.org/10.1371/journal.pbio.3000515 (visited on 04/03/2021).

[26] Joe Marcus. *Introduction to the Wright-Fisher Model*. Mar. 2016. URL: https://stephens999.github.io/fiveMinuteStats/wright_fisher_model.html (visited on 03/15/2021).

[27] Ignacio A. MartÃnez et al. "Engineered swift equilibration of a Brownian particle". en. In: *Nature Physics* 12.9 (Sept. 2016), pp. 843–846. ISSN: 1745-2473, 1745-2481. DOI: 10.1038/nphys3758. URL: http://www.nature.com/articles/nphys3758 (visited on 11/25/2020).

[28] Daniel W. McShea and Robert N. Brandon. *Biologyâs First Law*. URL: https://press.uchicago.edu/ucp/books/book/chicago/B/bo8642428.html (visited on 03/09/2021).

[29] Ville Mustonen and Michael LÃ€ssig. “From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation”. en. In: *Trends in Genetics* 25.3 (Mar. 2009), pp. 111–119. ISSN: 0168-9525. DOI: 10.1016/j.tig.2009.01.002. URL: http://www.sciencedirect.com/science/article/pii/S0168952509000250 (visited on 10/04/2020).

[30] Simona Negrini, Vassilis G. Gorgoulis, and Thanos D. Halazonetis. “Genomic instability â an evolving hallmark of cancer”. en. In: *Nature Reviews Molecular Cell Biology* 11.3 (Mar. 2010), pp. 220–228. ISSN: 1471-0072, 1471-0080. DOI: 10.1038/nrm2858. URL: http://www.nature.com/articles/nrm2858 (visited on 04/02/2021).

[31] Daniel Nichol et al. “Antibiotic collateral sensitivity is contingent on the repeatability of evolution”. en. In: *Nature Communications* 10.1 (Dec. 2019). ISSN: 2041-1723. DOI: 10.1038/s41467-018-08098-6. URL: http://www.nature.com/articles/s41467-018-08098-6 (visited on 09/21/2020).

[32] Daniel Nichol et al. “Steering Evolution with Sequential Therapy to Prevent the Emergence of Bacterial Antibiotic Resistance”. en. In: *PLOS Computational Biology* 11.9 (Sept. 2015). Ed. by Rustom Antia, e1004493. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004493. URL: https://dx.plos.org/10.1371/journal.pcbi.1004493 (visited on 09/05/2020).

[33] C. Brandon Ogbunugafor et al. “Adaptive Landscape by Environment Interactions Dictate Evolutionary Dynamics in Models of Drug Resistance”. en. In: *PLOS Computational Biology* 12.1 (Jan. 2016). Ed. by Carl T. Bergstrom, e1004710. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004710. URL: https://dx.plos.org/10.1371/journal.pcbi.1004710 (visited on 09/29/2020).

[34] Justin R. Pritchard, Douglas A. Lauffenburger, and Michael T. Hemann. “Understanding resistance to combination chemotherapy”. en. In: *Drug Resistance Updates* 15.5-6 (Oct. 2012), pp. 249–257. ISSN: 13687646. DOI: 10.1016/j.drup.2012.10.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S136876461200057X (visited on 04/03/2021).

[35] D A Ratkowsky et al. “Relationship between temperature and growth rate of bacterial cultures.” In: *Journal of Bacteriology* 149.1 (Jan. 1982), pp. 1–5. ISSN: 0021-9193. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC216584/ (visited on 10/19/2020).

[36] A. E. Siegman. “Simplified derivation of the FokkerâPlanck equation”. en. In: *American Journal of Physics* 47.6 (June 1979), pp. 545–547. ISSN: 0002-9505, 1943-2909. DOI: 10.1119/1.11783. URL: http://aapt.scitation.org/doi/10.1119/1.11783 (visited on 12/31/2020).

[37] Pauli Virtanen et al. “SciPy 1.0: fundamental algorithms for scientific computing in Python”. eng. In: *Nature Methods* 17.3 (Mar. 2020), pp. 261–272. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2.

[38] D. M. Weinreich. "Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins". en. In: *Science* 312.5770 (Apr. 2006), pp. 111–114. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1123539. URL: https://www.sciencemag.org/lookup/doi/10.1126/science.1123539 (visited on 07/10/2020).

[39] *WHO — Antimicrobial resistance: global report on surveillance 2014*. Publisher: World Health Organization. URL: http://www.who.int/drugresistance/documents/surveillancereport/en/ (visited on 04/02/2021).

[40] *WHO — WHO releases the 2019 AWaRe Classification Antibiotics*. Publisher: World Health Organization. URL: http://www.who.int/medicines/news/2019/WHO_releases2019AWaRe_classification_antibiotics/en/ (visited on 04/02/2021).

[41] S. Wright. "The roles of mutation, inbreeding, crossbreeding and selection in evolution". In: *Proc. Sixth Int. Congress on Genetics*. Vol. 1. University of Chicago Press, 1932, pp. 356–366.

[42] Boyang Zhao et al. "Exploiting Temporal Collateral Sensitivity in Tumor Clonal Evolution". en. In: *Cell* 165.1 (Mar. 2016), pp. 234–246. ISSN: 00928674. DOI: 10.1016/j.cell.2016.01.045. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867416300599 (visited on 04/03/2021).

[43] M H Zwietering et al. "Modeling of bacterial growth as a function of temperature." In: *Applied and Environmental Microbiology* 57.4 (Apr. 1991), pp. 1094–1101. ISSN: 0099-2240. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC182851/ (visited on 10/19/2020).