Dynamic Regulation of the Immune Repertoire of Bacteria

Zhi Zhang* and Sidhartha Goyal†
Department of Physics, University of Toronto.
(Dated: September 15, 2025)

The CRISPR-Cas system provides adaptive immunity in many bacteria and archaea by storing short fragments of viral DNA, known as spacers, in dedicated genomic arrays. A longstanding question in CRISPR-virus coevolution is the optimal number of spacers for each bacterium to maintain proper phage coverage. In this study, we investigate the optimal CRISPR memory size by combining steady-state immune models with dynamical antigenic traveling wave theory to obtain both analytic and numerical results of coevolutionary dynamics. We focus on two experimentally supported phenomena that shape immune dynamics: primed acquisition, where partial spacer-protospacer matches boost acquisition rates, and memory size fluctuations, where a short-term increase in memory size drives population dynamics. We find that under primed acquisition, longer optimal arrays benefit from maintaining multiple, partially matching spacers. In contrast, dynamic memory fluctuations favor shorter arrays by amplifying the fitness advantage of acquiring a few highly effective new spacers. Together, our results highlight that memory optimality is not fixed, but instead shaped by the interaction of acquisition dynamics and population-level immune pressures.

I. INTRODUCTION

Coexistence and defense against rapidly adapting viruses is a fundamental challenge across all domains of life, from bacteria to humans. While generalized innate immune systems provide broad protection against all viruses, they are likely not as effective as mechanisms that specifically learn and target particular pathogens. While the adaptive immune systems of vertebrates learns through the generation and selection of antibodies with high binding affinity to antigens, many bacteria and archaea species evolved an adaptive immune strategy through the CRISPR-Cas system, which specifically targets viral DNA [1–3]. While many different types of CRISPR-Cas mechanisms exist, they all rely on arrays of clustered regularly interspaced short palindromic repeats (CRISPR) loci. These loci consist of repetitive DNA sequences interspersed with short, unique sequences known as spacers, which are homologous to regions of previously encountered bacteriophage (phage) genomes. Immunity is achieved by transcribing each spacer into an RNA molecule (crRNA) to guide CRISPR-associated (Cas) proteins, which will cleave the recognized nucleic acids, thereby neutralizing the threat of foreign phage elements. The phage-derived DNA segments targeted by the CRISPR system before spacer integration are hereby referred to as protospacers.

The optimal number of spacers in each CRISPR locus remains a central question in the understanding of CRISPR-virus coevolution. Laboratory strains of bacteria [4, 5] often begin without spacers to acquire just a few spacers throughout the experiment. However, bacterial strains isolated from natural environments frequently have longer CRISPR cassettes, which may contain a few dozen to a few hundred spacers [5–7].

In this work, we assume that large-scale recombination occurs frequently enough within these populations to justify treating all spacers as part of a shared population pool. This regime may be particularly relevant in hot-spring microbial mats, where recombination contributes to genetic diversity on a scale comparable to mutation [8, 9] while the highly conserved spacer profiles suggest their CRISPR immune cycle operates on the order of years [10]. Our focus is on the long-term dynamics of CRISPR memory size in coevolving phage-bacteria systems under conditions of high horizontal gene transfer (HGT) rates, which are known to occur in hot-springs environments [11].

In such environments, prolonged coexistence between phage and host populations facilitates the advent of a coevolutionary steady-state, while HGT introduces variability in the effective memory size. Understanding how these forces interact to shape the optimality of CRISPR loci offers new insight into the diversity of immune strategies observed in natural microbial communities.

The optimal number of spacers in each CRISPR loci remains a central question in the understanding of CRISPR-virus coevolution. Laboratory strains of bacteria [4, 5] often begins without spacers to acquire just a few spacers throughout the experiment. However, bacteria strains isolated from natural environments frequently have longer CRISPR cassettes which may contain a few dozen to a few hundred spacers [5–7].

In this work, we assume that large-scale recombination occurs frequently enough within these populations to justify treating all spacers as part of a shared population pool. This regime may be particularly relevant in hot-spring microbial mats, where recombination contributes to genetic diversity on a scale comparable to mutation [8, 9] while the highly conserved spacer profiles suggest their CRISPR immune cycle operates on the order of years [10]. Our focus is on the long-term dynamics of CRISPR memory size in coevolving phage-bacteria systems under conditions of high horizontal gene trans-

* also at Department of Physics, University of Toronto.
† goyal@physics.utoronto.ca

fer (HGT) rates which are known to occur in hot-springs environments [11].

In such environments, prolonged coexistence between phage and host populations facilitates the advent of a coevolutionary steady-state, while HGT introduces variability in the effective memory size. Understanding how these forces interact to shape the optimality of CRISPR loci offers new insight into the diversity of immune strategies observed in natural microbial communities.

Previous models have demonstrated that increasing CRISPR-Cas memory size can lead to diminishing immune effectiveness due to intrinsic limitations in the number of interference complexes and the sensitivity of the CRISPR machinery [12]. Moreover, overexpression of CRISPR-Cas components has been associated with toxic side effects, particularly due to the risk of autoimmunity through self-targeting [13]. These arguments do not help to understand important phenomena in which cassette sizes are dynamic and change significantly in a bacterial population under constant attack from a rapidly evolving phage population. Here, we will focus on two phenomena: primed acquisition and memory size fluctuations through temporal regulation of spacer acquisition and loss rates.

Primed Acquisition. refers to a phenomenon where an already acquired primed spacer biases the acquisition probability of future protospacers, which generally leads to substantially increased acquisition rate [14, 15]. This effect has been observed across CRISPR subtypes and bacterial strains [16–19]. Accurately modeling primed acquisition has posed significant challenges, as it requires simulating bacterial populations carrying multiple spacers per individual alongside a highly diverse distribution of phage protospacers, while also incorporating acquisition biases informed by sequence similarity.

Memory Size Fluctuation is a result of the regulation of acquisition and loss rates. Regulation of CRISPR-Cas systems is fundamental since they are both metabolically expensive and pose a risk of autoimmunity [20]. In our model, we incorporate a tradeoff whereby temporary increases in acquisition efficiency are allowed at the cost of decreased net future spacer acquisition. Unlike previous models, which assume a fixed or average memory size across the population, we treat memory size as a dynamic, fluctuating property of each bacterium. This allows us to explore how variation around optimal memory size affects both bacterial and phage populations in a coevolutionary setting, which is an aspect that has not been addressed in previous models of CRISPR cassette dynamics.

While the joint dynamics of viral phage strains and CRISPR-Cas immunity have often been modeled using agent-based simulations that track individual bacterium and phage strains [21, 22], these models have been primarily used to study the population composition and evolutionary outcomes under phage pressure. We expanded the simulation models by moving away from agent-based frameworks, which focused on a more mechanistic description of immunity to a more spatial (antigenic) description of the dynamical landscape of bacteria–phage coevolution. Additionally, recent theoretical advances in adaptive immunity have highlighted the different modalities of coevolution between phage-bacteria, particularly how immune cross-reactivity leads to an influenza-like traveling wave regime [23]. In this work, we expand on this particular regime of high cross-reactivity by incorporating recent modeling approaches developed for vertebrate immune models to bacteria [24].

More specifically, we combine two complementary theoretical approaches to investigate optimal CRISPR memory size. First, we borrow from the statistical view of CRISPR–virus interactions, used to compute an optimal number of spacers under simplified biochemical assumptions, to compute a fitness function that controls the dynamics. Second, we used a time-dependent dynamical framework based on traveling wave theory to capture the coevolutionary immune dynamics in phage–bacteria systems. We demonstrate that primed acquisition tends to favor larger memory sizes, as individual spacers become less effective at interference, thereby requiring more redundancy. In contrast, memory size fluctuations tend to favor smaller memory sizes, as bacteria with fewer spacers can acquire new, more effective spacers, which can rapidly drive out phages in the environment.

II. MODEL AND METHODS

A. Adaptive Immunity Coevolution

While previous models used statistical physics arguments in population dynamics to solve for the optimal memory size under steady-state equilibrium, we use a new model which generates a dynamical fitness landscape between phage and bacteria within a joint antigenic space of dimension d. This model was initially proposed for the evolution of viruses through a shifting immunity mechanism [25–27]. As bacteria acquire spacers targeting current infectious phage protospacers, the fitness landscape experienced by the phage population shifts accordingly. Although phage protospacers mutate randomly in an isotropic manner, the selective pressure imposed by bacterial spacers drives phage evolution towards regions in antigenic space with lower immune coverage. This sets up a cyclical arms race, in which bacterial adaptation drives phage escape and vice versa. Unlike agent-based models that track individual entities, this framework captures the large-scale behavior of bacterial and phage populations by simplifying their interactions onto a shared antigenic landscape. While individual-level descriptions are lost, this model offers a powerful means to investigate the impact of larger memory size and other population-scale parameters while providing an independent population simulation scheme.

Following previous nomenclature and conventions from [26], we first assign positions (or "phenotypes") in a projected shared Euclidean antigenic space to both im-

mune spacers and viral protospacers. In the context of CRISPR-Cas systems, the interference machinery of CRISPR targets matching pairs of spacer and protospacer, which are DNA strands between 25-50 nucleotides in length. The Hamming distance, *i.e.* the number of mismatches between two spacers, is a natural metric in the antigenic space.

The coevolution dynamics are driven mainly by the coarse-grained phage population, whose population density in the antigenic space is denoted by $n(x,t)$. We can describe its evolution using a continuous mean-field equation that captures the effects of isotropic mutation and fitness-driven selection, where each infection cycle is assumed to occur on a timescale of bacterial generations (Fig. 1A).

$$\partial_t n(x,t) = D\partial_x^2\, n(x,t) + f(x,t)n(x,t) \tag{1}$$

After each infection cycle, individual phages in the population undergo mutation by randomly selecting a new location to jump to. This process results in an effective diffusion term with diffusivity $D = \mu\langle(\Delta x)^2\rangle/2$ where μ is the mean number of mutation events per cycle and $\langle(\Delta x)^2\rangle$ is the mean-squared jump length across the antigenic space. For simplicity, the average jump length of our system was chosen to be the smallest lattice unit length: $\Delta x = 1$. We believe that the fine-grained approximation of phage mutational effects in Eq. 1 is valid even though single substitutions in the protospacer sequence represent large jumps in our antigenic space. The fitness distribution of bacteria populations is much more fine-grained than their genetic distribution due to many non-genetic elements [28]. As such, to capture the evolution in the underlying population distribution on a generation scale, we overdefine the antigenic resolution.

The bacteria population, or more specifically the spacer population density, is denoted by $n_s(x,t)$. In this model, the bacteria population N_b is constant and not subject to phage selection. Events that disrupt coexistence, such as population runoff/extinction, will be described entirely by the change in the total phage population size $N(t) = \int n(x,t)dx$. The total spacer population size is given by $N_b M = \int n_s(x,t)dx$ where N_b is the fixed bacteria population and M is the total number of spacers. We fixed the expression level of the CRISPR system and assumed that all regulations that occur are averaged on a population scale. The evolution of the spacer population is governed by the acquisition and loss of spacers that occur on the same infection cycle time as that of the phages.

$$\partial_t n_s(x,t) = N_+(t)\, P_+(x,t) + N_-(t)\, P_-(x,t) \tag{2}$$

Here, the acquisition rates $N_+(t)$ and $N_-(t)$ are simply the total number of spacers acquired and lost at each infection cycle. These rates scale with the current phage population $N(t)$. The base case, without net memory gain or loss, is assumed to be $\alpha N(t)$ where α sets the per-cycle acquisition rate of protospacers. Throughout our simulations, this rate is usually set to $\alpha = 0.01$, which corresponds to the rate at which 1% of the total protospacer population will be converted to spacers at each timestep, *eg.* at 10 MOI and $\alpha = 0.01$, 10% of all bacteria will acquire a new spacer at each generation. While CRISPR acquisition rate varies widely depending on the species, subtypes, and environment [29, 30], this choice is motivated by typical phage-to-bacteria ratios (or multiplicity of infections, MOI) observed in natural systems [31] and coevolution experiments [4]. $P_+(x,t)$ and $P_-(x,t)$ are the acquisition probabilities of spacers mapped to point x. The unbiased acquisition probability is directly proportional to the phage density: $P_+(x,t) = n(x,t)/N(t)$. On the other hand, spacer loss is proportional to spacer abundance: $P_-(x,t) = n_s(x,t)/N_b\, M$ where we assume all spacers are forgotten with an equal rate. We do not assume a preferential ordering of spacers within a cassette; the older spacers will naturally fall off from the immune population as the two populations move within the antigenic space through each infection cycle.

Finally, the last key element to consider is the coverage $c(x,t)$ of the spacer population, which is the collective effect of all spacers on targeting phages across the antigenic space. This coverage function accounts for cross-reactivity, which is the ability of Cas interference proteins to partially recognize non-specific phage elements. To model the decrease in CRISPR effectiveness with mutational distance, we consider an exponential cross-reactivity kernel with radius r.

$$c(x,t) = \frac{1}{MN_b}\int e^{-\frac{|x-x'|}{r}}\, n_b(x',t)\, dx' \tag{3}$$

The choice for r in the cross-reactivity in Eq. 3 is motivated by several experiments where single point mutations in the protospacer sequence did not enable phage escape [32, 33]. Based on these experiments, we can set our cross-reactivity range to be around 1 point substitution difference in the spacer genetic sequence (Fig. 1B). The consequence of this scaling will be discussed further in section II C and reflects the anticipated range of adaptation speed and typical spread of protospacers we expect to observe, depending on the phage mutation rates.

B. Fitness Model

We choose a fitness function for the system that incorporates the fundamental constraints of the bacterial biochemical machinery. Previously proposed in [12] for phage and bacteria population in equilibrium, we model the probability of successful infection of a single bacterium (Eq. 4) as a function of the expected number of effective spacers in its CRISPR cassette (accounting for

cross-reactivity). The expected number of effective spacers is the previously mentioned coverage $c(x,t)$ of the spacer population. In our model, the infection probability decreases with the number of effective spacers (including the effects of cross-reactivity) present in each bacterium. However, even in the presence of a single effective spacer, a successful immune response is not guaranteed, as it will depend on the probability of expression and interference. For our purposes, we only consider the case where infection can occur only when one or fewer effective spacers are present. While a higher number of effective spacers can be considered, their contribution to the infection probability is very small in the relevant regime (SI 1C).

$$P_{\text{inf}}(x,t) = (1 - c(x,t))^M (1 + M \frac{c(x,t)}{1 - c(x,t)} Q) \quad (4)$$

Here, Q is the probability of unsuccessful interference despite the presence of a single effective spacer in the CRISPR cassette. Q is uniquely determined by the expression level of our CRISPR arrays, which depend on N_p and d_c. N_p, which is the number of expressed spacers into Cas-complexes. Furthermore, d_c is the number of bound complexes required to successfully interfere with the phage. Eq. 5 expresses Q as the probability that fewer than d_c complexes specific to the effective spacer are formed, given that each of the N_p complexes independently expresses that spacer with probability $1/M$.

$$Q = \sum_{d=0}^{d_c} \binom{N_p}{d} (\frac{1}{M})^d (1 - \frac{1}{M})^{N_p - d} \quad (5)$$

This infection probability effectively captures the tradeoff of memory size M compared to expression levels of N_p and sensitivity d_c. There exists an optimal memory size M that scales approximately as N_p/d_c, which minimizes P_{inf}. Memory sizes below this optimum do not generate enough coverage against phages, while higher memory sizes dilute the expression of each spacer, effectively reducing interference. Although the exact value of this optimality does depend on the specific coverage $c(x,t)$ from Eq. 4, this nonlinear effect was clearly observed in our dynamical simulations (Fig. 1C).

We adopt a simple fitness model in which the phage growth rate depends directly on the probability of successful infection. The key infection parameter is R_0, which is the basic reproduction number of the phage. As shown in various experiments [34], this value seems to be approximately 100 for various species of lytic phage. While more detailed models of fitness, such as SIR-type frameworks, have been used for both bacteria-phage and antigenic evolution [35, 36], we opt for a more coarse-grained description. Our focus is on the long-term coevolution of the system, allowing us to neglect the short-term fluctuations of subpopulations.

$$f(x,t) = \ln\left(R_0 \, P_{\text{inf}}(x,t)\right) \quad (6)$$

C. Antigenic Traveling Waves

Solving the general system of Eq. 1 and Eq. 2 analytically is typically impossible, but there exist some initial conditions for which the general long-term behavior of the population can be inferred. Such a non-trivial case is that of phage traveling waves, where the total phage population N remains constant and the population density $n(x,t) = n(x - vt) = n(u)$ is fixed along a co-moving frame with a steady velocity v. While many regimes of traveling waves exist, including the classical FKPP wave [37, 38], where single mutations are substantially beneficial and the resulting traveling wave is driven primarily by its most fit members. These regimes have vastly different behaviors when scaled to higher dimensions. Since a phage can undergo multiple distinct mutations to escape CRISPR targeting, the dimensionality of our system is at least $d \geq 2$. In this work, we will focus on the linear fitness regime as this system has been previously shown to produce long canalized solutions in 2D (Fig. 1B) [26]. These waves are characterized by the bacteria population forming a long tail of memory and are consistent with agent-based simulations of bacteria coevolution at large cross-creativity [23].

This linear fitness regime often arises when our cross-reactivity is much larger than the mutation step size: $\delta x \ll r$ [35], which motivates our choice of $r = 10^3 \delta x$ in the antigenic space. This regime also gives us an analytical approximation for the phage adaptation speed v and phage coevolution cluster spread σ^2 (Fig. 1D) [25]. When scaled appropriately to phage mutation rates per base pair which has been estimated to be from 10^{-8} to 10^{-6} substitution per nucleotide per infection cycle (s/n/c) for DNA phages and from 10^{-5} to 10^{-3} for RNA phages [39, 40] since some CRISPR subtypes can likely target RNA phages [41]. Our predicted adaptation speed matches the upper observed range of 10^{-4} to 10^{-3} s/n/c found in some phages [42]. The analytical expression for the traveling wave velocity v is given in Eq. 7 [25]. Here, the variable s represents our linear approximation to the fitness where $f(x,t) = f(u) \approx f_0 + su$.

$$v \approx (D)^{2/3} s^{1/3} (24 \ln(N (D s^2)^{1/3})^{1/3} \quad (7)$$

Finally, there are two conditions on the average coverage of the phage population by the spacer population, denoted by c^*. To ensure a consistent balance between spacer formation and phage escape and maintaining the long-term coevolution, c^* is first constrained by the wave adaptation speed and spacer turnover time $\tau = N_b M/(\alpha N)$. Furthermore, since the phage population must remain constant, the net fitness of the average phage must be 0: $f(c^*) = 0$. This imposes a new condition on c^*, which means it must also be the root of the fitness function defined in Eq. 6. The fitness can then be linearized around this average coverage: $s = \partial_u f(u) = \partial_c f(c^*) \, \partial_u c(u)$. $\partial_c f(c)$ can be found using Eq. 6 and 4.

$$\partial_c f(c) = \frac{M}{R_0} \frac{(1 + Q + (M-1)\frac{c}{1-c}Q)}{1 - c + McQ} \tag{8}$$

To close our traveling wave framework, we need to validate some final assumptions of the linear fitness regime. First, $v\delta t \ll r$ is maintained since adaptation at each infection cycle typically involves sub-nucleotide scale shifts, below the cross-reactivity threshold. Second, the memory of spacers extends beyond the current spread in the protospacer population or $\sigma \ll v\tau$. We believe it to be a reasonable assumption given data that shows long spacer persistence time (up to years) in bacteria with large CRISPR cassettes. Finally, we define a key lengthscale of $u_c = s\sigma^2/4D$ which represents the antigenic distance from wave center to most fit individuals, as those outliers play an important role in driving the adaptation [43].

III. RESULTS

A. Primed Acquisition

Primed acquisition is the phenomenon where new spacer acquisition is biased towards spacers that are genetically related to those already present in the cassette [44]. Priming can occur when a previously acquired spacer, even when it doesn't apparently confer any immune benefit, can serve to guide the CRISPR machinery to preferentially acquire spacers towards either a protospacer that's genetically similar or identical [17]. Although the precise mechanism of acquisition is still only partially understood, many experiments have been done for different CRISPR subsystems in different environments [16–19, 45]. In natural populations, we often see spacer pairs with highly similar genetic makeup following one another in an array (∼2% of arrays in E.Coli have repeated spacers according to [46]). Primed acquisition can also target protospacers located further on the phage genome. In some cases, the acquired protospacer could be genetically dissimilar while being acquired through a primed mechanism. Looking at the acquired protospacer mapping from several experiments [17, 18, 47], they suggest that priming does not appear to target any new genomic regions beyond those observed in naïve acquisition. Priming overall results in a narrower distribution in acquisition sites compared to naive acquisition.

Our mathematical model for primed acquisition first is the change in the protospacer acquisition probability $P_+(x,t)$ that will result in a bias towards the current immune (spacer) population. This is modeled in a similar way to the cross-reactivity kernel, where each spacer in the immune population will bias the phenotypical acquisition profile $\pi_+(x,t)$ by its own kernel (Fig. 2A).

$$\pi_+(x,t) = \int e^{-\beta|x-x'|} n_s(x',t)dx' \tag{9}$$

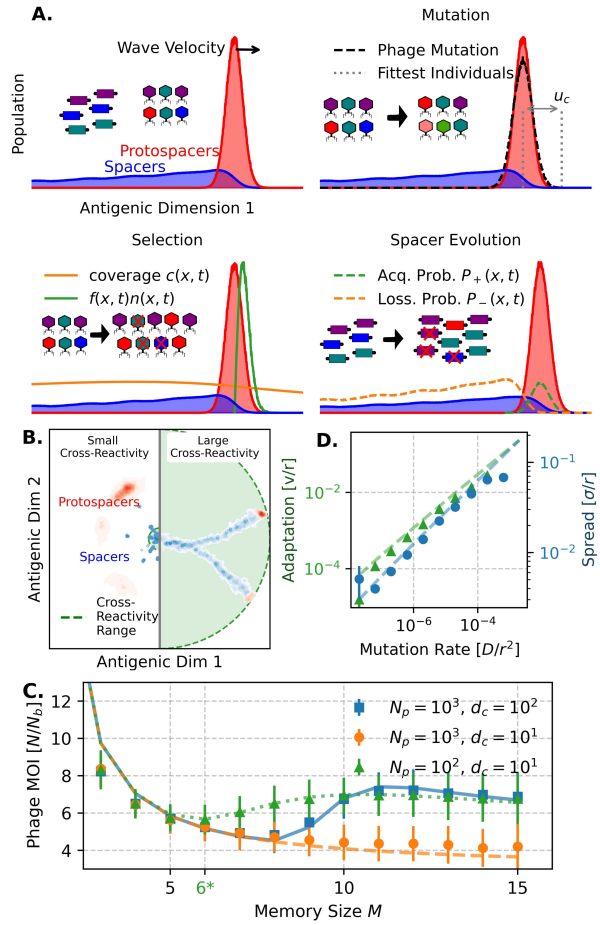

FIG. 1. Model for phage-bacteria coevolution captures biochemical limitation in a dynamically stable equilibrium. (A) Schematic of traveling linear fitness waves undergoing each simulation step. (B) Top-down view of traveling waves in 2 dimensions. Small cross-reactivity on the left leads to high rates of splitting, while large cross-reactivity on the right leads to canalized waves with long persistence time. (C) Dependence on M/Cas proteins on the amount of interference protein and binding sensitivity, there exists an optimal amount of spacers per bacterium that minimizes the coevolving phage population. Simulations were done with $R_0 = 100$, $\mu = 10$ and $\alpha = 0.01$. A linear correction was needed to correctly match simulation to theory, where $N_{\text{sim}} \approx 0.8 \cdot N_{\text{theory}}$. (D) Coexistence conditions for diversity and adaptation speed. Simulations were done with cross-reactivity $r = 1000$.

The acquisition profile for the whole spacer population depends on the multiplicity of each phenotype, depending on the current phage population: $P_+(x,t) \propto n(x,t)\,\pi_+(x,t)$ (Eq. 9).

Under the assumption that our phage population remains a Gaussian wave profile of the protospacer distribution in the direction of travel of the traveling wave, we obtain a new acquisition profile that's similarly a Gaussian distribution but now has an antigenic delay of $\beta\sigma^2$ in the traveling wave coordinate (Fig. 2A). It is important to note that this delay has an upper bound to the

expected distance to the fittest individuals $|\beta\sigma^2| < u_c$ since coexistence requires an overlap between the acquisition probability π_+ and the phage population distribution. Furthermore, within the previous assumption, it maintains all the previous constraints to maintain the analytical traceability of the linear fitness regime. While it may seem counterintuitive that individual spacers exhibit a weaker acquisition bias compared to their cross-reactivity, it is important to recognize that the primed parameter is an averaged effect over all possible acquisitions, including naive. Although the presence of a primed spacer biases acquisition, bacteria can still engage in naïve spacer acquisition. In contrast, cross-reactivity is an intrinsic phenomenon of CRISPR immunity that cannot be turned off.

Finally, we obtain a new coverage profile in front of the spacer distribution. While the coverage at the center of the phage wave should remain at c^* to maintain the zero growth condition. Priming induces a delay between the two waves (SI Section 4) and thus the coverage in front of the spacer wave must increase with priming. This leads to a different coverage at the leading edge of the spacer population:

$$c^*_\beta = c^* \, e^{\beta\sigma^2/r} \tag{10}$$

$$\frac{v\tau}{v\tau^*} \approx \frac{\alpha(\beta)}{\alpha(\beta=0)} = \frac{(1-c^*)e^{\beta\sigma^2/r}}{1-c^* e^{\beta\sigma^2/r}} \tag{11}$$

Then, using Eq. 9 and Eq. 10, we can get an estimate of the increase in phage MOI from priming (Eq. 11). This increase in MOI will also give us an estimate of the equivalent increase in acquisition rate necessary to balance out the evolutionary tradeoff of priming (Fig. 2C).

In Fig. 2A, unprimed acquisition is presented as a theoretical extension of the primed acquisition mechanism into the regime of negative $\beta < 0$. In this hypothetical scenario, bacterial immunity improves by preferentially acquiring spacers that are genetically farther away from the current spacer population. Although there is currently no experimental evidence for such behavior in CRISPR-Cas systems, this extension is conceptually interesting as it represents an alternative immune strategy, akin to predictive or exploratory immunity. While unprimed acquisition seems beneficial, its dependence on memory size seems weak (Fig. 2B).

The most striking result of our theory and simulations about primed acquisition is the immediate increase in phage MOI once priming is introduced (Fig. 2C). This indicates that across all memory sizes, the overall bacteria population performance worsens as the strength of the priming increases (assuming that acquisition rates remain identical). Nevertheless, bacterial systems with larger memory sizes tend to perform better than smaller memory sizes as they retain more information since new information becomes rarer and cross-reactivity more important. The advantage conferred by memory optimality

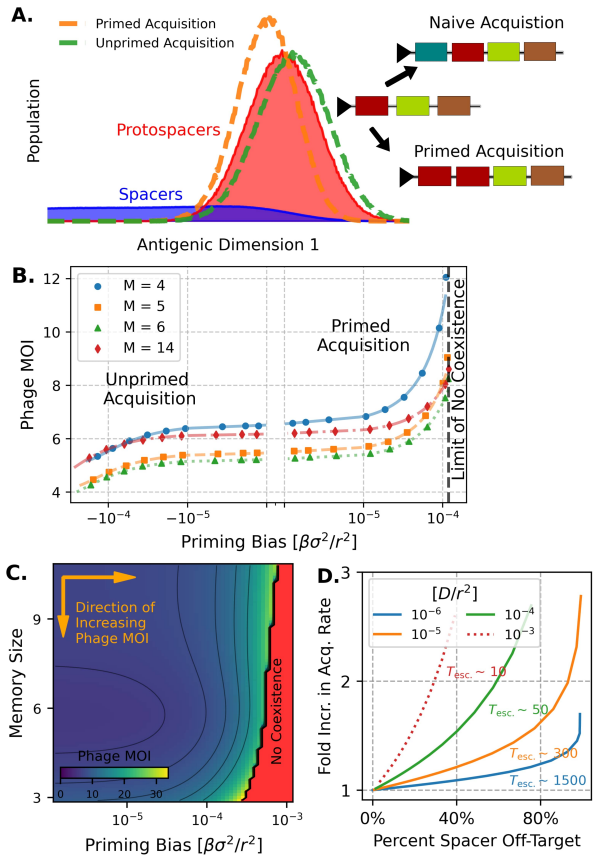

FIG. 2. Primed acquisition. (A) Acquisition profile of primed and unprimed acquisition. Primed acquisition induces a delay in the phage protospacer acquisition probability. Schematic of primed spacer acquisition: the new spacer acquired in the cassette is generically similar to another spacer already in the array. Unprimed acquisition is a theoretical extension of the primed case where bacteria try to predict future phage phenotypes. (B) Simulation results for different memory sizes with $N_p = 100$, $d_c = 10$, and $\mu = 10$. We found that simulation correctly matches our estimate with a correction factor of $\beta_{\text{sim}} \approx 4 \cdot \beta_{\text{theory}}$ (C) Coevolving Phage MOI vs priming strength, phage population increases as more and more of the newly acquired spacers are the result of priming. In the presence of priming, longer memory sizes become more beneficial. (D) Plots of acquisition and priming tradeoff. Increasing priming increases the perceived amount of spacers off-target compared to naive adaptation. As the mutation rate increases, the equivalent acquisition rate increases. Numerical analyses were done at $M = 3$ as lower memory sizes experience a larger impact due to priming.

also becomes less significant: the reduction in immune effectiveness caused by priming outweighs the benefits of memory retention.

Biologically, priming was thought to be beneficial because it increases the base acquisition rate of the bacterium for incoming primed spacers. While the underlying mechanisms have been described extensively, many subparts remain poorly understood. In our model, we

show that a mere threefold increase in acquisition rate is enough to counteract the effect of primed acquisition over the range of priming values (Fig. 2D). Note that the x-axis in Fig. 2D was chosen to give a more intuitive view of the priming strength: 0% indicates that all acquired spacers are indistinguishable from the naive case and 100% indicates that all spacers acquired would be different from the naive case, likely indicating that almost all spacers acquired had at least one duplicate already present in the population.

In multiple lab experiments, bacteria seem to result in a 100-1000-fold increase [17] while acquiring spacers with a primed target, leading us to believe that the biological benefit of faster acquisition outweighs the evolutionary trade-offs introduced by priming. The net increase in maximum effective priming strength at larger memory sizes (Fig. 3C) might also become more relevant if there is evidence that this increase in acquisition strength is biologically coupled with the priming factor per individual spacer. The net effect of priming is also predicted to be more visible in CRISPR bacteria coevolving with slower-mutating phages since the range of visible difference due to priming increases: a slower-evolving phage population leads to a smaller increase in primed MOI.

B. Memory Size Fluctuations

There are examples where the number of spacers can be transiently changed, but how that affects the co-evolutionary dynamics is not understood. For example, if CRISPR-Cas systems are encoded on mobile genetic elements [48], they can be dynamically regulated, *i.e.* these elements can be lost or gained by bacterial populations depending on their immediate environmental needs [49]. Why would bacteria turn off immunity? A key example of this involves horizontal gene transfers (HGT), which is any mechanism that leads to bacteria incorporating foreign genetic material. As such, HGT is thought to be incompatible with CRISPR-Cas immunity because of its function of defense against foreign genes. Indeed, there is experimental evidence that HGT events must be carried out while the adaptive immune system is suppressed or destroyed [50].

While there are different methods for modulating and/or disabling CRISPR-Cas activity, such as Cas gene deletion, transcriptome insertion, etc. [51]. These observations illustrate a broader evolutionary principle: adaptive immunity confers long-term benefits, but there may be short-term benefits of modulating immune memory size.

To model the effects of a single fluctuation event, the total number of bacteria N_b will remain constant, but the memory size of each bacterium will now undergo some fluctuations that will increase its memory size for a short duration (Fig. 3A). The idea is that bacteria will receive an excessive amount of new spacer for one generation while being evolutionary disadvantaged for a short time (about 100 generations in our model to match typical bacteria recovery time) before returning to steady state. Immediately following an increased acquisition event, the individual memory size of each bacterium $M(t)$ undergoes a recovery to its original or baseline memory capacity M_0, which we model as an exponential recovery with recovery time τ_m. The impact of each fluctuation event is governed by γ/α, which is the ratio between the increased acquisition compared to its baseline acquisition rate. Finally, for longer simulations with multiple events, we assume that those occur with a uniform time distribution with expected time between events to be longer than both the recovery timescale and cross-reactivity escape timescale: $T_{\text{HGT}} \gg \tau_m,\ T_{\text{escape}}$. The cross-reactivity escape timescale is the time required for the phage population to move beyond the range of cross-reactivity: $T_{\text{esc}} = v/r$.

$$M(t+\delta t) = \frac{\gamma N}{N_b}\delta(t) + M(t)(1 + \frac{1}{\tau_m}\left(1 - \frac{M(t)}{M_0}\right)) \quad (12)$$

To get an estimate of the effect of immediately adding spacers to the system we can estimate the transient phage population by averaging out the fitness effects $N(t) \approx \exp(\int \langle f(x,t)\rangle_x dt)N_0$ where N_0 is the original phage coexistence MOI and since we are in the linear fitness regime: $\langle f(x,t)\rangle_x \approx \partial_c f(c^*)\,|c_\delta(t) - c^*|$, we only need to consider the effect of the increased memory perturbation at the start of the event. The time of perturbation was taken to be $t = 0$ without loss of generality. We used the previously derived methods from antigenic traveling wave theory to find the appropriate expression for the net perturbation of coverage in time:

$$|c_\delta(t) - c^*| = \frac{\gamma\epsilon}{\alpha\tau}\exp(-|vt|/r) \quad (13)$$

The ϵ term in Eq. 13 is a correction term due to the non-linear nature of the fluctuation. Most phage phenotypes in the bulk will be extinct due to the reduced infectivity and leaving only the ones that are far away from the wave center. Finally, this leads us to a simple expression for the transient phage population (Eq. 14) with a noted dependence on memory optimality coming from $\partial_c f$ defined by Eq. 8. Analytical results for longer timescales can be done (SI section C). However, we will show that within our parameters, the phage population quickly returns to coexistence and does not depend on the specific recovery conditions.

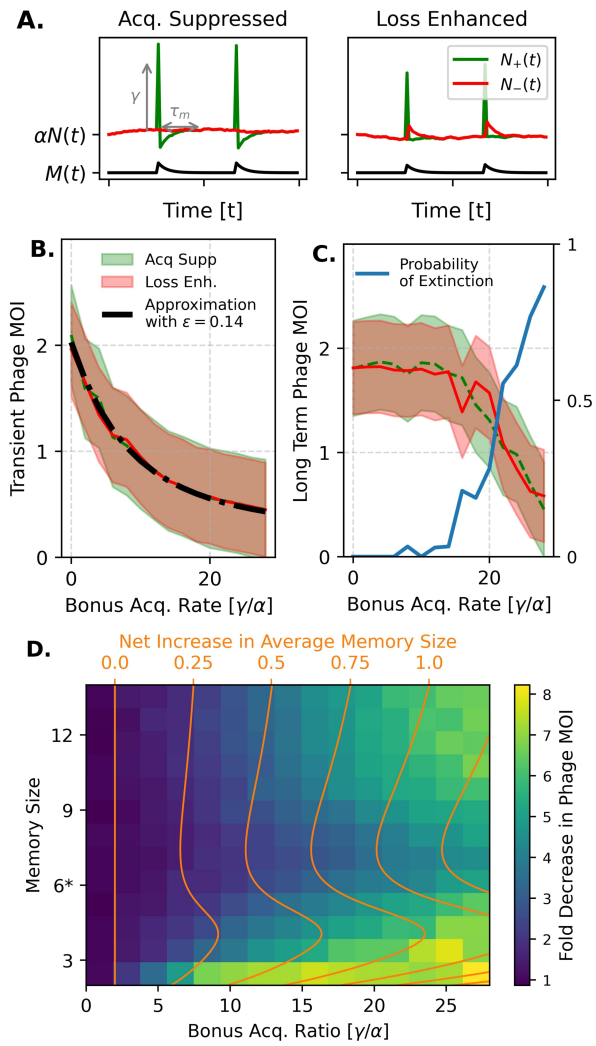

FIG. 3. Memory Fluctuation plots where the memory size of bacteria fluctuates with time (A) The two different schemes of memory fluctuations, where the tradeoff for short-term increase in memory size leads to a recovery time τ_m of either suppressed acquisition or increased spacer loss. (B) Simulations show that either scheme is equivalent in both the short and long term. Short-term phage MOI decreases accordingly with fluctuation size. Transient phage MOI were taken at $\tau_c = 20$. (C) Longer simulations were taken when $T_{\text{escape}} \sim 150$. As the fluctuation strength increases, the probability of extinction increases. (D) Fluctuation size depends on memory size; being at the optimal memory size (6*) does not lead to the most optimal decrease in phage MOI. The largest decrease in the transient MOI occurs when the memory size is small. Otherwise, the smallest increase occurs close to the optimal memory size. Simulations were done with $\alpha = 0.01$ and $\mu = 1$.

$$\frac{N_{\text{transient}}}{N_0} \propto \exp(-\gamma\,|\partial_c f(c^*)|) \tag{14}$$

There are two extreme forms of recovery dynamics we can impose on the system. Following an increased acquisition event, the number of spacers acquired must be either suppressed or the amount of spacers lost must be enhanced for the system to return to its initial state. This is modeled by either a suppressed amount of acquired spacers $N_+(t)$ (Eq. 16) or a time-dependent enhanced number of spacers lost $N_-(t)$ (Eq. 15 and Fig. 3A). Those two forms of recovery dynamics correspond to different biological mechanisms associated with immune suppression, more specifically, whether the acquisition mechanism or the interference mechanism is affected. In either case, the other corresponding rate is kept at $\alpha N(t)$ to ensure that only one component of the spacer turnover dynamics is perturbed for a given setup:

$$N_-(t) = \alpha N(1 - \frac{\gamma}{\alpha\tau_m} e^{-t/\tau_m}) \tag{15}$$

$$N_+(t) = \alpha N(1 + \frac{\gamma}{\alpha\tau_m} e^{-t/\tau_m}) \tag{16}$$

Overall, simulation results indicate that memory fluctuations are only relevant in the immediate transient timescale of evolution, such as the time-to-effect $\tau_c = u_c/v$, which is the timescale associated with when fitness effects catch up to the phage population at the front of the distribution. When averaged over a longer period of phage cross-reactivity escape time T_{escape}, the average phage population remains effectively unchanged regardless of the fluctuation scheme (Fig. 3B). This could be because, on average, even at fluctuation strengths that would drive out phages to extinction, average memory size only increased by around a single spacer. As a result, the new coexistence MOI of the phage is too close to the unperturbed number to meaningfully create an impact even if we had assumed those suppression to be permanent. This is especially evident in a low acquisition rate regime, where small changes in average memory size tend to drive phages to extinction before making evolutionary regime changes.

However, memory optimality significantly changes the impact of those memory fluctuations during transient phases. In contrast to primed acquisition, lower memory sizes actually perform better at reducing the phage population compared to larger memory sizes, although both extremes outperform memory sizes around the optimal value (Fig. 3C). While smaller cassettes are typically associated with higher baseline phage MOI, the net effect of adding a few spacers on average proves extremely beneficial. Conversely, while larger memory sizes also benefit from memory fluctuations, the net decrease is comparatively smaller than smaller memory sizes. This is because large memory sizes are already far from the optimal point of high biological tradeoffs, but will acquire significantly more phage protospacers due to their large coexisting population. A more analytic explanation of the overall memory dynamics would be that $|\partial_c f|$ from Eq. 8 is largest when M is kept small, thus leading to a larger decrease in Eq. 14. Ultimately, the most significant cost of memory optimality arises when memory sizes

are close/or slightly higher than optimal: they also represent the largest temporary increase in array length for a given increased acquisition rate, but the least amount of temporary decrease in phage population.

IV. DISCUSSION

We have shown that while optimal memory size in CRISPR-Cas systems is maintained in a dynamical equilibrium environment, different evolutionary tradeoffs emerge when phage and bacteria are considered together as dynamic variables. In a statistical setting, the limited number of Cas interference proteins and their binding sensitivity suggest the existence of an optimal cassette size. Under dynamical coevolution, this optimality shifts depending on the presence of primed acquisition and memory size fluctuations.

Primed acquisition tends to favor larger cassette sizes. Although individual spacers acquired through priming may confer weaker protection, this lowers the expected coverage of the spacer memories, which leads to an increased benefit from maintaining more memory, thereby pushing the optimal memory size higher. Furthermore, the substantially higher acquisition rate observed in biological systems compensates for the reduced specificity.

In contrast, memory fluctuations strongly benefit bacteria at smaller cassette sizes. When the goal is to drive phage populations to extinction rather than maintaining a smaller coexistence population, smaller increases in memory can yield a large selective difference for the bacterium. This is due in part to the fact that a small increase in coverage at low memory size can decrease the probability of infection substantially. This effect diminishes at larger memory sizes, where the relative benefit of an additional spacer becomes marginal.

In natural environments, bacteria usually have multiple cassettes in their genomes with different subtypes and cassette sizes [3]. While longer cassettes have been argued to better cover long-term coverage of common phage strains [12] and smaller memory size to rapidly acquire immune resistance to virulent strains [52]. Here, we suggest that the differences in evolutionary dynamics might also help explain the difference in cassette size depending on the eco-evolutionary constraints of the system. For phage-bacteria populations with longer coexistence time where phage and bacteria evolve gradually with one another, priming the spacer acquisition in exchange for increasing the acquisition is a logical evolutionary solution. In environments with highly diverse populations, competitive exclusion can be achieved more easily at the lower end of the memory sizes. Furthermore, the long-term cost of memory fluctuations seems to be minimal. Small fluctuations tend to revert to the traveling wave equilibrium, while large fluctuations are more likely to result in phage extinction. Thus, the system either self-restores or achieves its desired outcome depending on the fluctuation scale, without undermining long-term host fitness.

We have confirmed that the dependence of traveling wave dynamics on memory size holds both in one and two-dimensional systems. An important aspect not addressed in this work is the effect of priming and memory fluctuation on phage speciation. In higher-dimensional systems, the phage population can spontaneously split into multiple subpopulations, a process heavily dependent on cross-reactivity. We hypothesize that both priming and memory fluctuations could affect the rate and likelihood of such splitting. However, speciation was beyond the scope of this study as it is a separate measure not directly tied to memory size, which could be an avenue for future research.

Furthermore, while we assume that all the different tradeoffs associated with priming and fluctuations share unified dynamics, a more ecologically realistic analysis should include phages with varying infectivity. For instance, separate phage populations with different infectivity might lead to differing extinction probabilities with more meaningful ecological consequences. This would allow for a more nuanced analysis of how different memory strategies perform under diverse selective pressures, potentially revealing more meaningful aspects of the tradeoffs.

Additionally, another limitation of our traveling wave framework is that we neglected the effect of oscillations and other periodic dynamics in the underlying predator-prey model. Those oscillations play an important role in understanding the effect of population collapse and runoff, particularly in the case of large memory fluctuations. While we assume that under fluctuation, the dynamical regime of linear fitness remains valid during fluctuations, this may not hold in more extreme scenarios where alternative traveling wave solutions could become relevant. Indeed, more complex oscillatory behavior in CRISPR-virus coevolution has also been predicted to appear due to non-linear dependence of the host immunity on the virus load [36], which our current model does not capture. This limitation arises from the simplifying assumptions built into the total population dynamics of the model, which favors a stable dynamical regime in favor of a detailed spatial description of our populations (see Supplementary Information, Section 5C).

Finally, a key component of our framework assumes all spacers are shared collectively across the population. Notably, the traveling-wave regime at high cross-reactivity has also been observed in agent-based simulations that track individual spacer cassettes [22], indicating that this regime is not unique to our system. Nevertheless, the validity of this shared-memory approximation remains an open question and should be more critically evaluated by using empirical data from natural bacterial populations. While we hypothesize that our model is most applicable to populations in hot-spring communities, there is a strong lack of quantitative evidence to properly define the limits of our assumptions. Additionally, our dynamics are primarily driven by phages whose population dy-

namics remain difficult to characterize and notoriously hard to sequence.

Recent advances in single-cell sequencing and large-scale metagenomic analysis have provided unexpected insights into the communal nature of bacterial defense systems [53]. We hope that future experimental work leveraging these approaches can provide further insights into the role of horizontal gene transfer in spreading CRISPR elements and other immune components. Furthermore, a more detailed description of the interplay between individual and collective immunity could further extend our current simple model of memory size fluctuations.

CODE AVAILABILITY

Code for this paper is available at this link.

ACKNOWLEDGMENTS

This research was funded by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant and RGPIN-2021). We acknowledge helpful discussions and comments from all the members of the Goyal and Zilman groups.

[1] J. McGinn and L. A. Marraffini, Molecular mechanisms of CRISPR–Cas spacer acquisition, Nat Rev Microbiol **17**, 7 (2019), publisher: Nature Publishing Group.

[2] R. Barrangou and L. A. Marraffini, CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity, Molecular Cell **54**, 234 (2014), publisher: Elsevier.

[3] F. Tesson, R. Planel, A. Egorov, H. Georjon, H. Vaysset, B. Brancotte, B. Néron, E. Mordret, A. Bernheim, G. Atkinson, and J. Cury, A Comprehensive Resource for Exploring Antiphage Defense: DefenseFinder Webservice, Wiki and Databases, bioRxiv 10.1101/2024.01.25.577194 (2024).

[4] D. Paez-Espino, I. Sharon, W. Morovic, B. Stahl, B. C. Thomas, R. Barrangou, and J. F. Banfield, CRISPR Immunity Drives Rapid Phage Genome Evolution in Streptococcus thermophilus, mBio **6**, 10.1128/mbio.00262 (2015), publisher: American Society for Microbiology.

[5] C. Pourcel, M. Touchon, N. Villeriot, J.-P. Vernadet, D. Couvin, C. Toffano-Nioche, and G. Vergnaud, CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers, Nucleic Acids Res **48**, D535 (2020).

[6] Y. Agari, K. Sakamoto, M. Tamakoshi, T. Oshima, S. Kuramitsu, and A. Shinkai, Transcription Profile of *Thermus thermophilus* CRISPR Systems after Phage Infection, Journal of Molecular Biology **395**, 270 (2010).

[7] J. F. Heidelberg, W. C. Nelson, T. Schoenfeld, and D. Bhaya, Germ Warfare in a Microbial Mat Community: CRISPRs Provide Insights into the Co-Evolution of Host and Viral Genomes, PLOS ONE **4**, e4169 (2009), publisher: Public Library of Science.

[8] R. J. Whitaker, D. W. Grogan, and J. W. Taylor, Recombination shapes the natural population structure of the hyperthermophilic archaeon sulfolobus islandicus, **22**, 2354, publisher: Oxford Academic.

[9] M. C. Melendrez, E. D. Becraft, J. M. Wood, M. T. Olsen, D. A. Bryant, J. F. Heidelberg, D. B. Rusch, F. M. Cohan, and D. M. Ward, Recombination does not hinder formation or detection of ecological species of synechococcus inhabiting a hot spring cyanobacterial mat, **6**, 1540.

[10] M. Davison, T. J. Treangen, S. Koren, M. Pop, and D. Bhaya, Diversity in a polymicrobial community revealed by analysis of viromes, endolysins and CRISPR spacers, **11**, e0160574, publisher: Public Library of Science.

[11] G. Birzu, H. S. Muralidharan, D. Goudeau, R. R. Malmstrom, D. S. Fisher, and D. Bhaya, Hybridization breaks species barriers in long-term coevolution of a cyanobacterial population, eLife **12**, 10.7554/eLife.90849.1 (2023), publisher: eLife Sciences Publications Limited.

[12] S. Bradde, A. Nourmohammad, S. Goyal, and V. Balasubramanian, The size of the immune repertoire of bacteria, Proceedings of the National Academy of Sciences **117**, 5144 (2020), publisher: Proceedings of the National Academy of Sciences.

[13] S. Bradde, M. Vucelja, T. Teşileanu, and V. Balasubramanian, Dynamics of adaptive immunity against phage in bacterial populations, PLOS Computational Biology **13**, e1005486 (2017), publisher: Public Library of Science.

[14] T. J. Nicholson, S. A. Jackson, B. I. Croft, R. H. Staals, P. C. Fineran, and C. M. Brown, Bioinformatic evidence of widespread priming in type i and II CRISPR-cas systems, **16**, 566, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15476286.2018.1509662.

[15] S. H. Sternberg, H. Richter, E. Charpentier, and U. Qimron, Adaptation in CRISPR-Cas Systems, Molecular Cell **61**, 797 (2016), publisher: Elsevier.

[16] P. C. Fineran, M. J. H. Gerritzen, M. Suárez-Diez, T. Künne, J. Boekhorst, S. A. F. T. van Hijum, R. H. J. Staals, and S. J. J. Brouns, Degenerate target sites mediate rapid primed CRISPR adaptation, Proceedings of the National Academy of Sciences **111**, E1629 (2014), publisher: Proceedings of the National Academy of Sciences.

[17] R. H. J. Staals, S. A. Jackson, A. Biswas, S. J. J. Brouns, C. M. Brown, and P. C. Fineran, Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR–Cas system, Nat Commun **7**, 12853 (2016), number: 1 Publisher: Nature Publishing Group.

[18] O. Musharova, S. Medvedeva, E. Klimuk, N. M. Guzman, D. Titova, V. Zgoda, A. Shiriaeva, E. Semenova, K. Severinov, and E. Savitskaya, Prespacers formed during primed adaptation associate with the Cas1–Cas2 adaptation complex and the Cas3 interference nuclease–helicase, Proc. Natl. Acad. Sci. U.S.A. **118**, e2021291118 (2021).

[19] C. Richter, R. L. Dy, R. E. McKenzie, B. N. Watson, C. Taylor, J. T. Chang, M. B. McNeil, R. H. Staals, and P. C. Fineran, Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bidirectionally from the primed protospacer, Nucleic Acids

Res **42**, 8516 (2014).

[20] M. Zakrzewska and M. Burmistrz, Mechanisms regulating the CRISPR-cas systems, **14**, 1060337.

[21] E. V. Koonin and Y. I. Wolf, Evolution of the CRISPR-Cas adaptive immunity systems in prokaryotes: models and observations on virus–host coevolution, Mol Biosyst **11**, 20 (2015).

[22] M. Bonsma-Fisher, D. Soutière, and S. Goyal, How adaptive immunity constrains the composition and fate of large bacterial populations, Proceedings of the National Academy of Sciences **115**, E7462 (2018), publisher: Proceedings of the National Academy of Sciences.

[23] M. Bonsma-Fisher and S. Goyal, Dynamics of immune memory and learning in bacterial communities, eLife **12**, e81692 (2023), publisher: eLife Sciences Publications, Ltd.

[24] G. Altan-Bonnet, T. Mora, and A. M. Walczak, Quantitative immunology for physicists, Physics Reports Quantitative immunology for physicists, **849**, 1 (2020).

[25] L. S. Tsimring, H. Levine, and D. A. Kessler, RNA Virus Evolution via a Fitness-Space Model, Phys. Rev. Lett. **76**, 4440 (1996), publisher: American Physical Society.

[26] J. Marchi, M. Lässig, A. M. Walczak, and T. Mora, Antigenic waves of virus–immune coevolution, Proceedings of the National Academy of Sciences **118**, e2103398118 (2021), publisher: Proceedings of the National Academy of Sciences.

[27] L. Yan, R. A. Neher, and B. I. Shraiman, Phylodynamic theory of persistence, extinction and speciation of rapidly adapting pathogens, eLife **8**, e44205 (2019), publisher: eLife Sciences Publications, Ltd.

[28] A. J. Waite, N. W. Frankel, Y. S. Dufour, J. F. Johnston, J. Long, and T. Emonet, Nongenetic diversity modulates population performance, Molecular Systems Biology **12**, 895 (2016), https://www.embopress.org/doi/pdf/10.15252/msb.20167044.

[29] A.-N. Zhang, J. M. Gaston, P. Cárdenas, S. Zhao, X. Gu, and E. J. Alm, CRISPR-Cas spacer acquisition is a rare event in human gut microbiome, Cell Genomics **5**, 10.1016/j.xgen.2024.100725 (2025), publisher: Elsevier.

[30] L. J. Peach, H. Zhang, B. P. Weaver, and J. Q. Boedicker, Assessing spacer acquisition rates in E. coli type I-E CRISPR arrays, Front. Microbiol. **15**, 1498959 (2025), publisher: Frontiers.

[31] S. Batinovic, F. Wassef, S. A. Knowler, D. T. Rice, C. R. Stanton, J. Rose, J. Tucci, T. Nittami, A. Vinh, G. R. Drummond, C. G. Sobey, H. T. Chan, R. J. Seviour, S. Petrovski, and A. E. Franks, Bacteriophages in Natural and Artificial Environments, Pathogens **8**, 100 (2019).

[32] N. C. Pyenson, K. Gayvert, A. Varble, O. Elemento, and L. A. Marraffini, Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape, Cell Host Microbe **22**, 343 (2017).

[33] N. C. Pyenson and L. A. Marraffini, Co-evolution within structured bacterial communities results in multiple expansion of CRISPR loci and enhanced immunity, Elife **9**, e53078 (2020).

[34] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer, BioNumbers–the database of key numbers in molecular and cell biology, Nucleic Acids Res **38**, D750 (2010).

[35] V. Chardès, A. Mazzolini, T. Mora, and A. M. Walczak, Evolutionary stability of antigenically escaping viruses, Proceedings of the National Academy of Sciences **120**, e2307712120 (2023), publisher: Proceedings of the National Academy of Sciences.

[36] F. S. Berezovskaya, Y. I. Wolf, E. V. Koonin, and G. P. Karev, Pseudo-chaotic oscillations in CRISPR-virus coevolution predicted by bifurcation analysis, Biology Direct **9**, 13 (2014).

[37] R. A. Fisher, The Wave of Advance of Advantageous Genes, Annals of Eugenics **7**, 355 (1937), _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-1809.1937.tb02153.x.

[38] A. N. Kolmogorov, V. M. Tikhomirov, and V. M. Volosov, *Selected works of A.N. Kolmogorov* (Kluwer Academic Publishers, 1991).

[39] J. M. Cuevas, S. Duffy, and R. Sanjuán, Point Mutation Rate of Bacteriophage X174, Genetics **183**, 747 (2009).

[40] R. Sanjuán, M. R. Nebot, N. Chirico, L. M. Mansky, and R. Belshaw, Viral Mutation Rates, Journal of Virology **84**, 9733 (2010), publisher: American Society for Microbiology.

[41] K. S. Makarova, Y. I. Wolf, J. Iranzo, S. A. Shmakov, O. S. Alkhnbashi, S. J. J. Brouns, E. Charpentier, D. Cheng, D. H. Haft, P. Horvath, S. Moineau, F. J. M. Mojica, D. Scott, S. A. Shah, V. Siksnys, M. P. Terns, Venclovas, M. F. White, A. F. Yakunin, W. Yan, F. Zhang, R. A. Garrett, R. Backofen, J. van der Oost, R. Barrangou, and E. V. Koonin, Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants, Nat Rev Microbiol **18**, 67 (2020), publisher: Nature Publishing Group.

[42] H. A. Wichman, J. Millstein, and J. J. Bull, Adaptive Molecular Evolution for 13,000 Phage Generations, Genetics **170**, 19 (2005).

[43] R. A. Neher and O. Hallatschek, Genealogies of rapidly adapting populations, Proceedings of the National Academy of Sciences **110**, 437 (2013), publisher: Proceedings of the National Academy of Sciences.

[44] K. A. Datsenko, K. Pougach, A. Tikhonov, B. L. Wanner, K. Severinov, and E. Semenova, Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system, Nat Commun **3**, 945 (2012), publisher: Nature Publishing Group.

[45] S. A. Jackson, N. Birkholz, L. M. Malone, and P. C. Fineran, Imprecise Spacer Acquisition Generates CRISPR-Cas Immune Diversity through Primed Adaptation, Cell Host & Microbe **25**, 250 (2019), publisher: Elsevier.

[46] A. Kupczok, G. Landan, and T. Dagan, The Contribution of Genetic Recombination to CRISPR Array Evolution, Genome Biology and Evolution **7**, 1925 (2015).

[47] C. Rao, D. Chin, and A. W. Ensminger, Priming in a permissive type i-c CRISPR–cas system reveals distinct dynamics of spacer acquisition and loss, **23**, 1525, company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[48] N. D. McDonald, A. Regmi, D. P. Morreale, J. D. Borowski, and E. F. Boyd, CRISPR-Cas systems are present predominantly on mobile genetic elements in Vibrio species, BMC Genomics **20**, 105 (2019).

[49] J. C. Kosmopoulos, D. E. Campbell, R. J. Whitaker, and E. G. Wilbanks, Horizontal Gene Transfer and CRISPR Targeting Drive Phage-Bacterial Host Interactions and Coevolution in "Pink Berry" Marine Micro-

bial Aggregates, Applied and Environmental Microbiology 10.1128/aem.00177-23 (2023), publisher: American Society for Microbiology1752 N St., N.W., Washington, DC.

[50] L. A. Marraffini and E. J. Sontheimer, CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA, Science **322**, 1843 (2008).

[51] W. Jiang, I. Maniv, F. Arain, Y. Wang, B. R. Levin, and L. A. Marraffini, Dealing with the Evolutionary Downside of CRISPR Immunity: Bacteria and Beneficial Plasmids, PLOS Genetics **9**, e1003844 (2013), publisher: Public Library of Science.

[52] A. Martynov, K. Severinov, and I. Ispolatov, Optimal number of spacers in CRISPR arrays, PLOS Computational Biology **13**, e1005891 (2017), publisher: Public Library of Science.

[53] F. Tesson, A. Hervé, E. Mordret, M. Touchon, C. d'Humières, J. Cury, and A. Bernheim, Systematic and quantitative view of the antiviral arsenal of prokaryotes, Nat Commun **13**, 2561 (2022), publisher: Nature Publishing Group.