RICE UNIVERSITY

Physical Model of the Co-evolution of Bacteria and Viruses Mediated by CRISPR

by

Pu Han

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Michael W. Deem, Chair
John W. Cox Professor in Biochemical & Genetic Engineering and Professor of Physics and Astronomy

Ching-Hwa Kiang
Associate Professor of Physics and Astronomy and Bioengineering

Herbert Levine
Hasselmann Professor of Department of BioEngineering, Director, Center for Theoretical Biological Physics

Houston, Texas
June, 2016
ABSTRACT

Physical Model of the Co-evolution of Bacteria and Viruses Mediated by CRISPR

by

Pu Han

Bacteria and archaea have evolved an adaptive, heritable immune system that recognizes and protects against viruses or plasmids. This system, known as the CRISPR/Cas system, allows the host to recognize and incorporate short foreign DNA or RNA sequences, called ‘spacers’ into its CRISPR system. Spacers in the CRISPR system provide a record of the history of bacteria and phage coevolution. We use a physical model to study the dynamics of this coevolution as it evolves stochastically over time. We focus on the impact of mutation and recombination on the evolution and evasion of bacteria and phages. We discuss the effect of different spacer deletion mechanisms on the coevolutionary dynamics. We make predictions about bacteria and phage population growth, spacer diversity within the CRISPR locus, and spacer protection against the phage population.

An important feature of this coevolution is the multiple loci in the phages from which CRISPR may sample genetic material. We construct a model with multiple loci, in a two dimensional geometry. We show that to match recent experimental observations on the waves of replacement of spacers in CRISPR, multiple proto-spacer loci must be considered. With parameter values taken from the literature, the results are in agreement with experimental measurements of spacer abundance and turnover. We show that immune pressure upon the phages leads to greater rates of
phage evolution, by comparison to a neutral model, and we show a novel mechanism by which recombination in the phages leads to more effective phage escape from CRISPR recognition when there are multiple phage proto-spacer loci. We investigate the sensitivity of these results to model parameters, highlighting impacts on the immune pressure on the phages and to the evolvability of the phages.

The conditions for coexistence of bacteria and viruses is an interesting problem in evolutionary biology. In this thesis, we show an intriguing phase diagram of the virus extinction probability, which is more complex than that of the classical predator-prey model. As the CRISPR incorporates genetic material, viruses are under pressure to evolve to escape the recognition by CRISPR. When bacteria have a small rate of deleting spacers, this co-evolution leads to a non-trivial phase diagram of the virus extinction probability. For example, when the virus mutation rate is low, the virus extinction probability changes non-montonically with the bacterial exposure rate. The virus and bacteria co-evolution not only alters the virus extinction probability, but also changes the bacterial population structure. Additionally, we show that recombination is a successful strategy for viruses to escape from CRISPR recognition when viruses have multiple proto-spacers.
I would like to thank my advisor, Dr. Michael W. Deem, for his advices, guidance and continuous support to my study and research.

My sincere appreciation also go to my thesis committee: Dr. Ching-Hwa Kiang and Dr. Herbert Levine for their support and insightful suggestions on my thesis.

I am grateful to my current and former colleagues in Dr. Deem’s group: Man Chen, Liang Ren Niestemski, Jeong-Man Park, Yi Bao, Dong Wang, Andy Lee and Fengdan Ye. Without them, my research would not have been so enjoyable.

I would also like to thank my friends at Rice University and Houston for their help and moral support which made my life and study at Rice more delightful.

Finally, I would like to express my sincere gratitude to my parents and brother, for their love and encouragements. And my deepest thanks go to my lovely wife, Linglin Yu, for her love, care and support.
Contents

Abstract ii
Acknowledgments iv
List of Figures viii
List of Tables xii

1 Introduction 1
1.1 Introduction to CRISPR .............................. 1
   1.1.1 CRISPR .................................... 1
   1.1.2 CRISPR is an immune system ..................... 2
   1.1.3 Mechanism of CRISPR action ........................ 3
   1.1.4 CRISPR maintenance ............................. 5
   1.1.5 Bacteria/Phages co-evolution .................... 6
1.2 Motivation ............................................ 7
1.3 Previous Publications ................................. 8

2 Mean Field Model of the Co-evolution of Bacteria and Viruses 11
2.1 Methods ............................................. 11
   2.1.1 Co-evolution model ............................... 11
   2.1.2 Events .......................................... 12
   2.1.3 CRISPR details ................................... 13
   2.1.4 Phage details .................................... 13
   2.1.5 Mean field approximation and Monte Carlo method . 14
# Table of Contents

2.1.6 Density dependence of growth rate ........................................ 16
2.1.7 Method validation ......................................................... 19
2.1.8 Diversity measurements .................................................. 19
2.1.9 Spacer effectiveness ....................................................... 20
2.1.10 Recombination in the Phages .......................................... 21

2.2 Results ................................................................................ 21

2.2.1 Diversity versus position .................................................. 22
2.2.2 Protection versus position .................................................. 22
2.2.3 Deletion mechanism .......................................................... 24
2.2.4 Recombination versus mutation .......................................... 25

2.3 Discussion ............................................................................ 27

2.4 Conclusion ............................................................................ 32

2.5 Table of Parameters ............................................................. 33

3 Waves of Spacer Replacement in CRISPRs Acquired from Phages with Multiple Proto-spacers 35

3.1 Methods .............................................................................. 35

3.1.1 Events ............................................................................. 36
3.1.2 Parameters and initial conditions ......................................... 40

3.2 Results ................................................................................. 40

3.2.1 Daily waves of replacement in the CRISPR spacers ........... 40
3.2.2 Spacer turnover ................................................................. 41
3.2.3 Sensitivity to parameter values .......................................... 42
3.2.4 Neutral model ................................................................. 49
3.2.5 CRISPR accelerates the co-evolution ............................... 53

3.3 Discussion ............................................................................ 53

3.3.1 Spacer distribution for different phage diffusion rates ........ 56
3.3.2 Spacer turnover ................................................................. 56
### 3.3.3 Immunity ........................................... 57
### 3.3.4 Peak in phage population .............................. 58

### 3.4 Conclusion ............................................. 58

### 3.5 Table of Parameters .................................... 60

### 3.6 Supporting Information .................................. 60
  - 3.6.1 Neutral model ........................................ 60
  - 3.6.2 Phage diffusion ....................................... 62

### 4 Nonclassical Phase Diagram for Virus Bacterial Co-evolution

Mediated by CRISPR .......................... 76

- 4.1 Methods .............................................. 76
- 4.2 Results .............................................. 78
- 4.3 Discussion ............................................ 85
- 4.4 Appendix .............................................. 85
  - 4.4.1 Table of parameters ................................. 85
  - 4.4.2 Master equation ..................................... 87
  - 4.4.3 Mean field equations ............................... 90
  - 4.4.4 Number of spacers .................................. 91
  - 4.4.5 Number of proto-spacers ........................... 91

### 5 Conclusion .............................................. 93

Bibliography ............................................ 96
# List of Figures

1.1 Illustration of CRISPR. ........................................ 4
1.2 Addition of a new spacer to the CRISPR locus at the leader-proximal end. ........................................ 6

2.1 In the Markov process, five categories of transition events change the state of the system. ........................................ 17
2.2 The solution of the differential equation and the result of the stochastic method. ........................................ 19
2.3 The recombination procedure of phages. .............................. 22
2.4 Population of bacteria and phages with time for constant and density-dependent phage growth rates. ........................................ 23
2.5 Diversity of spacers at CRISPR position \( i \) at different times. .... 24
2.6 Protection afforded by spacers at different positions of CRISPR at different times. ........................................ 25
2.7 Spacer diversity versus location when the probability of deleting spacer \( i \) is proportional to \( i \). ........................................ 26
2.8 Diversity at different positions of CRISPR at different times for \( l = 2 \) with mutation only and recombination only. ........................................ 28
2.9 Diversity of the phage for \( l = 2 \) with mutation only. ........... 28
2.10 Protection at different positions of CRISPR at different times for \( l = 2 \) with mutation only and recombination only. ........................................ 29
2.11 Immunity at different mutation rate and recombination rate when \( l = 1 \). 30
3.1 Bacteria infection events. ........................................ 38
3.2 The phages spread out after lysis. .............................. 38
3.3 Different initial geometric conditions for infected bacteria. 41
3.4 Daily abundance of the most 10 abundant spacers and the spacer turnover during the first 10 days. .......................... 42
3.5 Daily abundance of the 10-most-abundant spacers. .......... 43
3.6 The population of phages for one proto-spacer locus and 30 proto-spacer loci. ............................................ 44
3.7 The population of phages for different mutation rates of phages . . 45
3.8 The distribution of phages and bacteria. ....................... 46
3.9 The correlation of healthy and infected bacteria as a function of radius. 47
3.10 The correlation of healthy and infected bacteria as a function of radius for different infection rates. ..................... 47
3.11 The population of phages when phages only had recombination or only had point mutation. .............................. 48
3.12 Recombination versus point mutation for phage population and infection probability. ................................... 50
3.13 The difference in phages not recognized by CRISPR for recombination or point mutation dynamics, $\langle S_{\text{recombination}} \rangle - \langle S_{\text{point mutation}} \rangle$, versus the number of initial viral strains. .............................. 51
3.14 A neutral model of the diversity of spacers. .................... 51
3.15 The diversity of spacers versus time when CRISPR was active and inactive. .................................................. 53
3.16 The diversity of phages versus time when CRISPR was active and inactive. .................................................. 54
3.17 Illustration of the diffusion process. ........................... 63
3.18 In the neutral model, when the population of bacteria was very large, the difference between the spacer’s diversity in different positions was almost 0. ........................................ 64
3.19 Daily abundance of the 10-most-abundant spacers. .................. 65
3.20 Sensitivity of spacer turnover to different number of proto-spacers per phage. ........................................ 65
3.21 Sensitivity of spacer abundance to different number of proto-spacers per virus. ........................................ 66
3.22 Spacer turnover for different probabilities of acquiring new spacers. .. 67
3.23 Spacer turnover for different mutation rates of proto-spacers. .... 67
3.24 The population of phages for different probabilities of acquiring a new spacer. ........................................ 68
3.25 The fraction of bacteria with spacers in the total bacteria population. 69
3.26 The population of bacteria when there was one proto-spacer locus and 30 proto-spacer loci. ...................................... 69
3.27 The population of bacteria and phages when there was one spacer or four spacers. .......................................... 70
3.28 The population of bacteria and phages when there was one spacer or 30 spacers. ........................................ 70
3.29 The difference in phages not recognized by CRISPR for recombination or point mutation dynamics, \( \langle S_{\text{recombination}} \rangle - \langle S_{\text{point mutation}} \rangle \) for different probabilities of acquiring new spacers. ......................... 71
3.30 The difference in phages not recognized by CRISPR for recombination or point mutation dynamics, \( \langle S_{\text{recombination}} \rangle - \langle S_{\text{point mutation}} \rangle \) for different proto-spacer mutation rate. .............................. 71
3.31 Different of the infection probability, \( S \), between the recombination and mutation for different probabilities of acquiring new spacers and mutation or recombination rates. ............................... 72
3.32 Spacer turnover for different phage decay rates. ..................... 73
3.33 The comparison of Fig. 3.5 and Fig. 1A. .......................... 73
3.34 Daily abundance of the 10-most-abundant spacers and spacer
turnover during the first 10 days. ....................................... 74
3.35 Frequency of newly acquired spacers types in the position +2 in the
first 10 days. .............................................................. 74
3.36 Daily abundance of the most 10 abundant spacers and spacer
turnover during the first 10 days. ....................................... 75

4.1 The extinction probability of bacteria and phages. .................... 79
4.2 The extinction probabilities of phages and bacteria for different \( \gamma \) and \( d \). 82
4.3 The diversity of the first spacer. ........................................ 83
4.4 The extinction probability of phages and bacteria for point mutation
and recombination. ......................................................... 84
4.5 The average number of spacers in CRISPR. .......................... 91
4.6 The average number of spacers in CRISPR when \( \mu = 10^{-8} \). ....... 92
4.7 The extinction probability of bacteria and phages when \( n_p = 1 \). ... 92
List of Tables

2.1 Table of parameters .............................................. 34

3.1 Table of parameters .............................................. 60

4.1 Table of parameters .............................................. 86
Chapter 1

Introduction

1.1 Introduction to CRISPR

1.1.1 CRISPR

The newly discovered CRISPR system in bacteria and archaea is a fascinating system for experimentalists and theorists. CRISPR was initially discovered in the gene sequence of *Escherichia coli* [1]. Ishino et al. found an unusual structure in the 3’-end flanking region of the *iap* gene, namely repeats of the same 29 nucleotides, each separated by spacers of 32 non-repeatable nucleotides. Inside each repeat there are two short sequences of DNA that are nearly reverse complements of each other, *i.e.* nearly palindromic sequence. These two palindromic DNA sequences, *e.g.* TTGTAC and GTACAA in Fig. 1.1a, are transcribed into RNA sequences *e.g.* UUGUAC and GUACAA in Fig. 1.1b, that can base pair to form a stable hair pin loop, as in Fig. 1.1c. The discovery of these repeat sequences in *E. coli* spawned an extensive search for similar interspersed and repetitive DNA sequences in other bacteria and archaea. To date these structures have been identified in 40% of bacteria and 90% of archaea [2, 3, 4, 5]. They are now termed clustered regularly interspaced short palindromic repeats (CRISPR) [6].

Bacteria or archaea can encode one or more CRISPR systems in their genome. Although the CRISPR gene structure varies greatly between different species, it has a few common features. It always has a leader-repeat-spacer-repeat-spacer... organ-
nization. Repeats are the regions with the same nucleotide sequences with nearly palindromic symmetry. The length of the repeat ranges from 23 bp to 47 bp in different organisms. Spacers are the nucleotide regions between the repeats. The length of the spacer ranges from 21 bp to 72 bp in different organisms [3]. Leaders are AT rich sequences located at the 5’ end of the CRISPR system. Leaders serve as an indicator of the beginning of the CRISPR system and give it a polarity. Leader sequences of different CRISPR systems in the same species are the same [6, 7]. Leaders serve as the recognition site for the addition of new spacers, and new spacers are always added to the leader proximal end of the CRISPR [8, 9].

1.1.2 CRISPR is an immune system

CRISPR is part of the immune system of bacterial and archaea. This functionality was discovered while studying phage resistance in *Streptococcus thermophilus*, a lactic acid bacterium used in the production of yogurt from milk [10, 11, 12]. Like other types of bacteria, milk lactic acid bacteria can be infected by viruses known as bacteriophage, and phage infection is the major cause of milk fermentation failure. It was observed that not all of the milk lactic acid bacteria cease to grow upon challenge with bacteriophage [9, 13]. Some of the bacteria were phage resistant. Extensive genome sequencing of lactic acid bacteria and virulent phage led to a better understanding of the difference between phage resistant bacteria and phage sensitive bacteria. Bacteria with phage resistance have CRISPR systems in their genome with spacer sequences that match the specific phage to which they are resistant. Upon challenge with a new type of phage, vulnerable bacteria strains have the ability to acquire sequences, termed protospacers, from the phage genome that are inserted into their CRISPR next to the leader sequence. This newly acquired spacer contains genetic information
from the population of currently infecting phage. Descendants of these bacteria inherit their ancestor’s genome with the inserted spacers and are phage resistant. The connection between the CRISPR related immune system, and phage resistance was confirmed in several laboratory experiments [8]. A phage-resistant strain of bacteria remained phage resistant upon removal of all spacers except the one derived from the phage of interest. A phage resistant strain became phage sensitive upon removal of the specific spacer derived from the phage of interest even when all the other non-relevant spacers were present. When the matching spacer was added back into the CRISPR, the previously phage-sensitive bacteria became resistant to that specific phage, but were susceptible to new phages with different genomes. Bacteria can be immune to more than one type of phage if different CRISPR spacers match different phage genomes. To sum up, a CRISPR spacer matching a protospacer sequence in a phage genome provides resistance against that specific phage.

1.1.3 Mechanism of CRISPR action

The CRISPR system as well as the CRISPR associated (Cas) genes located in the vicinity of CRISPR are essential components of a functional CRISPR-associated complex of antiviral defense (Cascade) system. The components of the Cascade system can be isolated from bacteria and reconstituted for study in vitro, and the mechanism of Cascade action has been investigated in both *E. coli* and *S. thermophilus* [14, 15, 8, 16]. The Cascade defense process starts with spacer acquisition, proceeds with CRISPR expression, and finishes with CRISPR interference. During spacer acquisition, Cascade recognizes a foreign nucleic acid sequence, *i.e.* a protospacer. In the case of *S. thermophilus*, short conserved regions within a few bases of the protospacer sequence are identified as CRISPR motifs. These CRISPR motifs serve as signal
Figure 1.1: a) Typical repeat in S. thermophilus CRISPR1, data taken from [7]. b) RNA repeat. c) Left: secondary structure (hairpin) of RNA repeat. Right: pre-crRNA. d) Targeting and protecting during CRISPR immunity.

for the bacteria’s Cascade system to recognize the protospacer. Upon recognition, a new sequence of nucleic acid identical to the protospacer is generated and integrated into the CRISPR system as a newly acquired spacer. During CRISPR expression, the spacer is transcribed into pre-crRNA. With the participation of the Cas protein, pre-crRNA matures into small crRNAs [14]. Within each crRNA, the transcript of a single palindromic repeat folds into a stable hairpin shape termed a handle with several stable base pairs, which may serve as a platform for RNA-binding Cas proteins [7]. The CRISPR spacer is connected to one end of the hairpin structure in each crRNA transcript. The crRNA is transported to the target phage DNA [17, 18, 19] or RNA [20]. During CRISPR expression, the crRNA guide the Cas complex to foreign nucleic acids. The expressed spacer sequence provided by the crRNA is thought to recognize and guide the complex to bind the specific protospacer target sequence.
Cas proteins with nuclease activity then cleave the invading nucleic acids in order to inhibit phage infection. If there is no match between the CRISPR spacer and the phage DNA, the phage is not neutralized by the spacer transcript. In this case, the phage can reproduce inside the bacteria and lead to bacteria lysis and death. This process is illustrated in Fig. 1.1d.

1.1.4 CRISPR maintenance

CRISPR can acquire new spacers from protospacer sequences within a phage genome. This process is shown in Fig 1.2. Although the addition of a single new spacer is a low probability event, it can occur in at least some cells within a population of bacteria upon the phage challenge. Newly acquired spacers are inserted at the leader-proximal end [7, 16]. The number of repeat-spacer units per CRISPR ranges from a few to hundreds [19, 21], with a typical length of 30–100 spacers. For example, one strain of *S. thermophilus* has 32 spacers [9]. Since the CRISPR cannot grow to infinite length, deletion of older spacers is required, and deletion has been documented concomitantly with spacer addition. The mechanism of spacer deletion is unclear, especially the location of deletion. One hypothesis is that the oldest spacer is the least needed in the current viral environment and should be deleted. Under this hypothesis, the deletion always happens at the leader distal end of the CRISPR. Another hypothesis is that the deletions can happen in the middle of CRISPR locus at random locations, or at locations following a certain distribution, such as linear distribution within the cluster, perhaps by recombination. Deletion at the leader-distal end and internal deletion are found in bacteria [22]. Moreover, deletion of greater than one spacer at a time is also observed in *S. thermophilus* [9]. Taken altogether, these experiments indicate that maintenance of CRISPR system by spacer addition and deletion occurs.
1.1.5 Bacteria/Phages co-evolution

The CRISPR immune system imposes a selection pressure on the phages. Conversely, the phages also impose a selection pressure on the bacteria. The efficiency of the CRISPR immune system has a direct impact on the fitness of the bacteria. Since bacteria are surrounded by multiple strains of phages, bacteria with CRISPR-containing spacers matching many phages are more likely to survive and reproduce. Due to evolution of the phages, the CRISPR spacers must be continually updated to protect against new phage sequences. The phage fitness depends on its ability to avoid recognition by CRISPR. There are several mechanisms of phage evolution. Phages can overcome CRISPR recognition by acquiring a single mutation [8, 9]. This will cause a mismatch between the spacer transcript of crRNA and the invading phage protospacer, leading to failure of the CRISPR interference. Another CRISPR-evading strategy is recombination between phages during coinfection in bacteria. Recombination is an inherent feature of phage evolution. Metagenomic studies of different phage popula-
tion document large scale recombination in phage [23]. Recombination can lead to a more rapid rate of phage evolution away from CRISPR recognition than does point mutation. First, recombination of previously established mutations incorporates mutations that have already been selected for increased fitness, i.e. mutations at less risk of altering essential protein function. Second, recombination can integrate multiple beneficial point mutations in one step, and it may be the case that \( l > 1 \) mismatches between the crRNA and protospacer are required for the phage to escape CRISPR recognition. Thus, we expect recombination will allow phage to evade CRISPR more effectively than point mutation alone.

## 1.2 Motivation

The role of recombination has been under-studied, even though it is a significant driver of evolution [19]. Recombination and other mechanisms for generating genomic diversity are especially important in coevolving systems with large population density. Here, we study the effect of recombination on bacteria-phage coevolution. We use a physical model that incorporates selection pressure, and we allow both recombination and mutation to occur. The theory and model provide time-resolved ‘snapshots’ of this coexistence. We focus on the case where bacteria and phage coexist, i.e. neither bacteria nor phage are driven extinct. That is, we choose parameters of the system to establish a robust coexistence so that bacteria and phage both have stable populations without species extinction. We study the effect of different spacer deletion mechanisms, which is important for understanding how CRISPR functions and interpreting patterns of CRISPR variation in natural populations of bacteria.
1.3 Previous Publications

The discovery of the CRISPR system pushed forward our understanding about the evolutionary dynamics of bacteria and phages. Several models were established to explain the intriguing features of CRISPR and the coevolution of prokaryotes and phages. The attempt to explain the coexistence and coevolution of bacteria and phages has a long history. Early models focused on ecological effects [24] and the “killing” the winner hypothesis [25].

An early theoretical treatment introduced a mean field model and a stochastic Markov process to describe the population dynamics of bacteria and phages. The result of this model showed that the spacers near the leader end are more diverse than that of the “older” spacers [26]. In Chapter 2 of this thesis, an extension to a density dependent phage growth rate showed similar results after an initial transient [27]. In contrast to our previous works, the model in Chapter 3 differs in three respects: detailed consideration of the phage infection, latent time, and burst size; multiple proto-spacers; and spatial effects. An eco-evolutionary model of CRISPR with imperfect immunity also showed that both bacteria and phages were highly diversified by coevolution and that diversity decreased with distance from leader [28]. In a follow-up paper [29], a metric of population-wide distributed immunity (PDI) was defined to quantify the immunity distributed among the host-viral population. This model showed that the number of viral proto-spacers, mutation rate, host spacer acquisition rate, and spacer number could change the host-viral population structure by a distributed immunity. The dynamics of the distributed immunity of this model agreed with the patterns observed in experimental data [30].

An ecological model addressed the question of why and how CRISPR establishes and is maintained in a bacterial population [31]. In this model, bacteria could have
immunity other than the CRISPR system. The results indicated that in the presence of other immunity to phages, the conditions that maintain CRISPR in a bacteria population are narrower than those without presence of other immunity.

A spatial model of the bacteria and phages interaction showed that to reach the coexisting state of bacteria and phages, there should be enough phage, but not so many that they overwhelm the CRISPR immunity, in the system [32]. In a follow-up paper, the fitness cost of spacers was taken into consideration [33]. Due to the spatial inhomogeneity and the fitness cost of spacers, it was observed that evolution favors an intermediate number of spacers.

A “population-genetic” model [34] reported the gradual loss of bacterial diversity through selective sweeps in the host population. The bacteria that gained spacers targeting several different viral sub-populations arose rapidly in the host population, resulting in selective sweeps. This model also showed that the leader-distal spacers of the array was conserved even though the old spacers did not provide immunity against current phages. Increasing the spacer deletion rate repressed the bacterial immunity and led to a viral bloom.

Why CRISPRs are more common in archeae than in bacteria has been examined [35]. One hypothesis is that the viral mutation rate is higher in the mesophilic environments, leading to the higher abundance of CRISPRs in archaea. A stochastic model of viral-CRISPR coevolution was used to test this hypothesis. In the model, bacteria had innate immunity, and CRISPR had a fixed cost. The model showed that a decreased viral mutation rate increases the prevalence of CRISPR in archae, and CRISPR appeared only at an intermediate level of innate immunity. In a follow-up paper [36], a model with explicit population dynamics showed that CRISPR was ineffective for extremely large populations. Because mesophiles usually have larger
population sizes, this model gave another explanation for the increased prevalence of CRISPR in hyperthermophiles compared to mesophiles.

A probabilistic model for three CRISPR spacer content evolution mechanisms was presented in [37]. The parameters in the model were estimated by a maximum likelihood method. Three *Yersinia pestis* data sets were analyzed, and this model captured some characteristics of the data sets, such as the average diversity of spacers.

CRISPR-Cas immunity reflects a balance between suppression of phage infection and acquisition of beneficial mobile genetic materials [38, 39]. A model that considers conjugation of beneficial plasmids shows that plasmids may be more likely to evade CRISPR-Cas immunity by inactivation of functional CRISPR-Cas rather than by mutation of the target proto-spacers [40].

Recent experiments have shown that waves of phage proto-spacers are sequentially incorporated into a bacterial CRISPR system [41]. In this system there are multiple proto-spacer loci in the phage genome, and replacement of a CRISPR spacer can occur by incorporation of a proto-spacer from any of the phage loci. Interestingly, in this experiment there is only a single spacer locus in the bacteria, despite the multiple proto-spacer loci in the phages. A common limitation of many previous theoretical models is the assumption of a single phage proto-spacer locus, with some notable exceptions [28, 35, 36]. We here show that multiple proto-spacer loci are required to explain recent experimental data [41]. As has been noted [32], spatial heterogeneity is also an important feature favoring bacteria and phages coexistence and diversity. We here develop the multiple proto-spacer model in two dimensions to explore the spatial effect on the evolutionary dynamics of CRISPR and phages.
Chapter 2

Mean Field Model of the Co-evolution of Bacteria and Viruses

In this chapter, we establish a physical model of bacteria-phage convolution and study the impact of recombination, spacer deletion, and non-linear growth dynamics in this system. In the Methods section, we introduce mean field population dynamics equations and a stochastic simulation to sample the underlying Markov process. In the Results section, we discuss the effect of a non-linear density-independent growth rate on the population dynamics. We demonstrate heterogeneity of spacer diversity in different spacer locations within CRISPR. We predict the spacer usage with respect to spacer location. We investigate three different deletion mechanisms: deletion of the oldest spacer, deletion of older spacers with increased probability, and deletion of a random spacer. We study the impact of recombination and mutation on the evolutionary dynamics. In the Discussion session, we relate observations from our physical model to experimental and natural coevolution of bacteria and phage. We conclude in the last section.

2.1 Methods

2.1.1 Co-evolution model

We consider a coexisting system composed of bacteria and phages. Each bacterium can have a different CRISPR system, and there are multiple strains of phages. The
evolution of the bacteria and phages is interrelated and changes dynamically. Bacteria with higher fitness have more descendants, and the number of bacteria with that CRISPR system increases. At the same time, phages with unsuccessful CRISPR evading strategies cease to reproduce, and the number of those strains decreases. Since there are different sequences of bacteria and strains of phages, the total population of both the bacteria and phages can reach a steady state even though the population of each bacterial sequence or phage strain may be changing with time. This steady state is what we are interested in. The total number of bacteria and the total number of phages reach the maximum steady-state values, after an initial exponential growth phase.

2.1.2 Events

We describe the bacteria-phage community dynamics using a population dynamics model [26]. The population structure of the bacteria and phages changes based on several events. The bacteria can reproduce at a certain rate until they reach the maximum capacity, defined by the available resources. This rate can be constant or dependent on the phage population. The phages reproduce at another rate, which can also be constant or dependent on the total number of bacteria, until the maximum carrying capacity of the phages is reached. Upon exposure to phage, a bacterium has the opportunity to acquire a protospacer from that phage, which will be inserted into the leader-proximal end of CRISPR in that bacterium.

We assume that phages can mutate at some defined rate or they can recombine with other phages, also at a certain rate. Either process leads to avoidance of CRISPR recognition by the evolved phages. Recombination has the additional advantage that it can combine the benefits of multiple mutations, which can provide the recombined
phage with a higher fitness.

2.1.3 CRISPR details

The number of spacers that a CRISPR contains varies between types of bacteria. Most CRISPR contain fewer than 50 spacer repeats. We here set the maximum number of spacers in CRISPR to be 30. Upon phage attack, a new spacer can be acquired and inserted at the leader-proximal end of CRISPR. We label the leader-proximal position of the spacer to be position 1, and the leader-distal position of spacer to be position 30. In general, positions with smaller index host “younger” spacers. When a new spacer is added to a CRISPR that already has 30 spacers, spacer deletion occurs to maintain a maximum length of CRISPR. We investigate three different types of spacer deletion: delete the “oldest” spacer; delete a spacer with a possibility proportional to its distance to the leader end, and delete spacer at a random position.

2.1.4 Phage details

Each phage genome is assumed to contain only a single protospacer. This protospacer is what our dynamics depend on, so we track only the protospacer part of the phage genome. It is known that a single phage may contain multiple protospacers [8], often localized in the early expressed, coding region of the phage genome [9]. We here simplify the biology, assuming only a single protospacer per phage. The protospacer in the genome of each phage is expressed as a bit string. Each bit of the string can be either “0” or “1”. The length of the phage bit strings is \( n \), and there are \( 2^n \) types of phage genomes. In our simulation, we set \( n = 10 \). Therefore, we have \( 2^{10} \) genome types available for phage. Initially, the population distribution of phage follows a logarithmic distribution \( p(i) = \log(150) - \log(i), i = 1, \ldots, 150 \), where \( p(i) \) is
proportional to the percentage of \(i^{th}\) phage strain. This distribution has been used to fit experimental data \[42\]. We start with 149 strains of phage with this distribution, and they evolve over time.

When phage replicate, there is a chance for phage to mutate with a rate \(\mu\) per genome per replication. This is part of the CRISPR-evading strategy of phage. We choose a random location in the phage genome to be the location of the point mutation. As this location, we alter the phage sequence from “1” to “0” or from “0” to “1”. The probability for \(n\) mutations in one sequence is \(\mu^n e^{-\mu}/n!\).

2.1.5 Mean field approximation and Monte Carlo method

We used two methods to study this system: a standard numerical fourth-order Runge-Kutta method to solve the mean-field differential equations and a stochastic simulation using the Lebowitz-Gillespie algorithm \[43\] to sample the Markov process. Both methods converge to the same result in the limit of a large population.

In the mean-field or infinite-population, spatially homogeneous limit the system can be described by the differential equations \[26\]

\[
\frac{dx_{i,j}}{dt} = \left[ cx_{i,j} - \beta \sum_{k \neq i,j} v_k x_{i,j} + \beta \gamma \sum_m x_{j,m} v_i - \beta \gamma \sum_k x_{i,j} v_k \right] \Theta \left( x_{\text{max}} - \sum_{i,j} x_{i,j} \right) \tag{2.1}
\]

\[
\frac{dv_k}{dt} = \left[ rv_k - \beta \sum_{i,j} x_{i,j} v_k (\delta_{i,k} + \delta_{j,k}) \right] \Theta \left( v_{\text{max}} - \sum_k v_k \right) \tag{2.2}
\]

We have set the number of spacers in CRISPR to 2 initially. We also initially do not consider virus evolution. The population of bacteria with spacer \(i\) in position 1 and spacer \(j\) in position 2 is \(x_{i,j}\), where the maximum bacteria population is \(x_{\text{max}}\). The phage population is \(v_k\), where the maximum phage population is \(v_{\text{max}}\). Here \(\Theta \left( x_{\text{max}} - \sum_{i,j} x_{i,j} \right)\) is a step function. When \(x_{\text{max}} > \sum_{i,j} x_{i,j}\), \(\Theta\) has a value of 1, otherwise it is 0. Each population grows until it reaches its maximum value.
Maximum population sizes are given from ecology and are due to maximal carrying capacities in the case of bacteria or number of available hosts in the case of viruses. The dynamics of the population depends on the events described earlier. Bacteria grow at a rate $c$ until they reach the maximum population. Phage grow at a rate $r$, which could be dependent or independent of bacteria population $\sum_{i,j} x_{i,j}$. Bacteria have an exposure rate $\beta$ to the phage. Upon phage attack, bacteria have a probability $\gamma$ of acquiring a new spacer from the protoscaler in the phage genome, independent of the current spacers within the CRISPR. The rate of spacer addition is $\beta \gamma \sum v_k$. Conversely, when the bacterial CRISPR system is unable to recognize the invading foreign genetic material, lysis of the bacteria occurs after infection. The rate of bacteria killed by phage is $\beta \sum_{k \neq i,j} v_k$. The term $\beta \gamma \sum_{m} x_{j,m} v_i$ represents the process of converting other types of bacteria into type $i,j$. The term $\beta \gamma \sum_{k} x_{i,j} v_k$ represents the process of converting type $i,j$ into other types of bacteria. When the CRISPR locus contains spacers matching the viral genetic profile, the phage is disarmed and eliminated. The rate of phage killed by bacteria is $\beta \sum_{i,j} x_{i,j} (\delta_{i,k} + \delta_{j,k})$.

We also use the Lebowitz-Gillespie algorithm to sample the stochastic process of bacteria phage coevolution. In this stochastic process we include the mutation and recombination events described in Section 2.4. The Lebowitz-Gillespie algorithm computes trajectories for a Markov process in which the rate $\phi_i$ of the every event $i$ is known. The algorithm works as follows: at time $t = 0$ a list of all possible rates $\phi_i$ in the system is formed. One event is randomly chosen to happen from the list with a probability proportional to its rate. There are five categories of events in the list. 1) A bacteria can be randomly chosen to reproduce at a rate $c$. Overall, bacteria reproduce at a rate of $\phi_1 = cx$, where $x = \sum_{i,j} x_{i,j}$ is the total bacteria population and $x_{i,j}$ is the population of bacteria strain with spacers $i$ and $j$. 2) A
bacteria can be killed by phage with a rate of $\beta \sum_{k \neq i,j} v_k$, where $v_k$ is the population of phage strain with protospacer $k$. Overall, bacteria are killed by phage at a rate of $\phi_2 = \sum_{i,j} \beta \sum_{k \neq i,j} v_k x_{i,j}$. 3) A new spacer can be added to a randomly chosen bacteria with a rate of $\beta \gamma v$, where $v = \sum_k v_k$ is the total phage population. This new spacer is chosen from the protospacers among all the phage according to the rate $\beta \gamma v_k$. Overall, new spacers can be added to bacteria with the rate $\phi_3 = \beta \gamma v x$. 4) A phage can be randomly chosen to reproduce at a rate of $r_0$. Overall, phage reproduce at a rate of $\phi_4 = rv$. 5) A phage can be killed by bacteria at a rate of $\beta \sum_{i,j} x_{i,j} (\delta_{i,k} + \delta_{j,k})$. Overall, phage are killed by bacteria at a rate of $\phi_5 = \sum_k \beta \sum_{i,j} v_k x_{i,j} (\delta_{i,k} + \delta_{j,k})$. Time is incremented by $-\ln(u)/\sum \phi_i$, where $u$ is a uniform random number $\in (0, 1]$. The rates of all possible events are then updated, if they have changed. We iterate this process until time reaches the specified final time. See Fig. 2.1. When the maximum population size is reached and a growth move is attempted, a random member of the population is deleted during replication.

Initially, we start with 149 types of phage with a logarithmic-decay population structure, $p(i) = \log(150) - \log(i), i = 1, \ldots, 150$, where $p(i)$ is proportional to the percentage of $i$th phage strain. Every new bacteria has a CRISPR with 30 empty spacers, i.e. all value of spacers are null. The initial population of phage is 1000, the initial population of bacteria is 4000.

### 2.1.6 Density dependence of growth rate

The model of the previous section applies when the phage grow not only in the bacteria we study, but also in another set of background bacteria. These background bacteria are the hosts providing the approximately constant growth rate of the phage, $r_0$. The populations of bacteria and phage are dynamically changing with time. At
Figure 2.1: In the Markov process, five categories of transition events change the state of the system. The rates of all of these events are denoted by $\phi_i$. Processes 2, 3, and 5 all result from phages infecting the bacteria. Processes 4 results from phage infecting some bacteria, which could be the population under study, or a different host population of bacteria. There is an additional category of events, not shown in this figure, which is evolution of the virus due to mutation or recombination.

short times, starting from an initially small population, the bacteria grow exponentially until stabilizing at the maximum population size. This is shown by the magenta curve overlaid on top of the red curve in Fig. 2.4. We now set the maximum number of spacers to be 30. Similarly, the phage population grows exponentially for a short period of time until stabilizing at the maximum phage population size. If the background bacteria are quite numerous, then the phage can have a growth rate independent of the bacteria under study, labeled by $x$.

If there is no such set of background bacterial hosts, the phage growth rate may depend directly on the bacteria we study, labeled by $x$. In this case, the reproduction rate of phage $k$ is a time dependent function of the bacteria population, i.e. $r_k = r_0 \sum_{i \neq k, j \neq k} x_{i,j} / x_{\text{max}}$, where $x_{\text{max}}$ is the maximum population of bacteria. The average replication rate is $\langle r_k \rangle = \sum_k \sum_{i \neq k, j \neq k} x_{i,j} v_k / (x_{\text{max}} \sum_i v_i)$. A simplified form if most
of the bacteria population is available to any given phage is $r = r_0 \sum_{i,j} x_{i,j}/x_{\text{max}}$ for all phages. At short times, the immunity has not built up yet, and the condition $i \neq k, j \neq k$ is irrelevant. The only difference between the non-linear and constant growth rates at short time is a slightly slower increase of the phage population in the non-linear case. The blue and black curves in Fig. 2.4 show phage populations with constant and density-dependent growth rates differ only at short times.

In general, we are interested in the case where the phage and bacteria populations reach steady-state. In this case, there will be an effective growth rate of the phage. This effective growth rate is $r_0$ in the linear model. Because the bacteria reach the maximum population size quickly, and because the non-linear growth model is different from the linear model only when the bacteria are below the maximum population size, the growth dynamics of the non-linear and linear model differ only at short times. At long times, most of the bacteria population will still be available to any given phage for growth, and so $\langle r_k \rangle \approx r \approx r_0$. In particular, we find $\langle r_k \rangle / r$ is unity for $t < 200$, and rises only to 0.97 at $t = 2000$ for the parameters we use in section 3.2. Even for the parameters corresponding to a more effective immune system in section 3.4, this quantity is unity for $t < 200$, 0.93 for $t = 600$, and 0.82 for $t = 2000$. These results justify the assumption that most of the bacteria population is available to growth of any given phage. Thus, the non-linear and linear growth models only differ at very short times when the bacterial population is not yet the maximum size, or at rather long times if the diversity of the phage population is driven to low values. The non-linear and linear growth models would also differ if the bacterial population were driven extinct, a situation we do not consider in this chapter.
Figure 2.2: The solution of the differential equation and the result of the stochastic method. Parameters are $c = 0.15$, $r = 0.045$, $\beta = 2 \times 10^{-5}$, $\gamma = 0.1$, $v_{\text{max}} = 17500$, and $x_{\text{max}} = 4500$. There are 2 spacers in a CRISPR. The error bars are one standard error. The bacterial growth rate sets the explicit time scale in this model.

### 2.1.7 Method validation

To validate the stochastic method, we compare it to the solution of the differential equation. Both results converge to the same result in the limit of large population. This convergence is evident in the population versus time curve, Fig. 2.2.

### 2.1.8 Diversity measurements

The Shannon entropy of spacers at a specific location is a measure of the diversity of the spacers at that location. A larger Shannon entropy indicates more diversity of spacers. We, thus, measure the diversity of spacers by the Shannon entropy: The diversity for the $i^{th}$ spacer is defined as

$$D_i = - \sum_k p_i(k) \ln p_i(k)$$  \hspace{1cm} (2.3)
where $p_i(k)$ is the observed probability to have sequence $k$ at position $i$.

Because new spacers are always added to the leader-proximal end, the spacer with smaller index is “younger” than the spacer with larger index. If the phage do not impose selection pressure on the bacteria, all spacers are randomly selected and inserted into the leader-proximal end of the CRISPR, and we will observe homogeneous diversity at all positions of CRISPR. With selection pressure, the diversity of CRISPR may decline toward the leader-distal end of the CRISPR if the distribution of phage genotypes is biased, as has been observed in experiments [7, 8, 44, 23, 42]. Gaps and insertions in the CRISPR array may result in different bacteria having nearly the same spacer content, but at slightly shifted spacer positions. This dephasing will be observed as a reduction of $D_i$ values relative to what could be observed with multiply aligned sequences.

We also define a diversity that averages out the effect of spacer position. The definition is

$$D = -\sum_k \left[ \sum_i \frac{p_i(k)}{N} \right] \ln \left[ \sum_i \frac{p_i(k)}{N} \right]$$

(2.4)

where $N$ is the number of spacers within the CRISPR.

In addition, we define the diversity of the phage. This is simply Eq. (2.3), but applied to the single protospacer in each phage, rather than spacer $i$ of each bacteria.

2.1.9 Spacer effectiveness

We count the number of matches between the spacer at position $i$ of the CRISPR and the current phage strains. We use this count as a measure of protection offered by spacer. The bigger this count is, the more frequently the spacer can be used, and the more effective the spacer is at protecting the bacteria from the phages. A
higher frequency of usage indicates a strong protection. Since the spacers at the leader-proximal end are recently acquired and reflect the current viral environment, these spacers should be highly used and offer the strongest protection against current phages. Within one CRISPR, we expect a decline of the protection with respect to position from the leader-proximal.

2.1.10 Recombination in the Phages

Another CRISPR-evading strategy of phage is recombination. Recombination can recombine multiple existing point mutations or even different strains. When two phages infecting the same cell recombine, they swap a portion of their genetic materials. This swapping is a random process. For a given sequence, recombination happens with a randomly chosen other sequence at a given probability $\nu$ per sequence per replication. In this way, mutation rates and recombination rates have the same units. Also with this definition, equal recombination and mutation rates imply an equal probability of changing a given sequence by an evolution event, except for the relatively rare occurrence of recombination between two identical sequences. We simulate this random process as follows: To assemble the recombined phage from two existing parental strains, we first pick randomly which sequence to start copying, then the copy process continues on that sequence with probability $1 - p_c$ or switches to the other with probability $p_c$ until an entire offspring sequence is created. See Fig. 2.3.

2.2 Results

We are interested in the coexistence of bacteria and phage at long times. In these models, both the total phage and bacteria population grow to their maximum carrying capacity at long times. Density-dependent and constant growth rates produce the
Figure 2.3: When multiple phages infect the same bacteria, two parents may produce a descendant by the polymerase copying along one strand with probability \(1 - p_c\) and switching to another strand with probability \(p_c\). This process leads to recombination between the phage genomes.

2.2.1 Diversity versus position

The diversity of spacers in the CRISPR system is measured using Shannon entropy, Eq. 2.3. We keep track of spacer diversity with respect to the position of the spacer. This is shown in Fig. 2.5. The diversity of the spacers at the leader-proximal end is higher than the diversity of the spacer at the leader-distal end.

2.2.2 Protection versus position

We define protection as a measure of spacer effectiveness, \(i.e.\) match of CRISPR spacers to phages. We calculate the ability of spacers at position \(i\) to protect against the current viral population. This is shown in Fig. 2.6. Since bacteria have the ability to acquire new protospacers from the phage population, and the insertion of new spacers happens at the leader-proximal end of the CRISPR, it is expected that
Figure 2.4: Population of bacteria and phages with time. We show results for constant and density-dependent phage growth rates. The parameters are $c = 0.15$, $r = 0.05$, $\beta = 2 \times 10^{-5}$, and $\gamma = 0.1$. The mutation rate per sequence per replication is $\mu = 0.01$. The maximum population of phage is $v_{\text{max}} = 6000$, and the maximum population of bacteria is $x_{\text{max}} = 12000$. The maximum number of spacers in a CRISPR is 30. When the number of spacers in the CRISPR array is over 30, the oldest spacer is deleted from the leader-distal end. There are 149 phage strains with a logarithmic initial population distribution.
the spacers at the leader-proximal end have the highest frequency of usage. The protection of the spacer falls off rapidly with distance from leader. Nonetheless, due to the random loss of spacers, some infection memory can be lost as time elapses.

2.2.3 Deletion mechanism

The diversity of spacers with respect to the location of spacers for three different deletion mechanisms is shown in Figs. 2.5 and 2.7. There is a small but significant difference in the dynamics of these three models. Diversity versus time shows the same trend for all three methods: the leader-proximal end is more diverse, and leader-distal end is less diverse. Although the diversity decreases toward the leader-distal end of the CRISPR, it decreases the least when the oldest spacer is deleted. In random deletion, every spacer in CRISPR has the same possibility to be deleted. Even the newer spacer, closer to the leader-proximal end, can be deleted. As a result, the
Figure 2.6: Protection afforded by spacers at different positions of CRISPR at different times. Protection is defined as the number matches between the spacer and the protospacers in the current phage population. The parameters are the same as in Fig. 2.4.

decrease of the diversity at the leader-distal end from the leader-proximal end is the largest among the three for random deletion. The decrease of diversity for the linear deletion mechanism is midway between that for the other two deletion mechanisms because leader-distal spacers with less diversity have more possibility to be deleted.

2.2.4 Recombination versus mutation

We compare the impact of mutation and recombination on phage evolution. We define the minimum number of mismatches between the CRISPR and phage required for the phage to escape recognition as $l$. The bigger the value of $l$, the harder it is for the phage to escape from the targeting spacer. A value of $l = 1$ means if there is one or greater mismatch between the spacer and the phage genome, the spacer provides no protection against the phage. A value of $l = 2$ means that CRISPR recognizes the
Figure 2.7: Spacer diversity versus location when the probability of deleting spacer $i$ is proportional to $i$. a) When the number of spacers in the CRISPR array is over 30, one spacer is selected to be deleted with a possibility proportional to its distance to the leader proximal end. b) When the number of spacers in the CRISPR array is over 30, one spacer at a random location is deleted. The parameters are the same as in Fig. 2.4.

Phage even if there is one mismatch between the spacer and the phage genome. For $l = 2$, if the spacer and the phage differ at one position, the phage is still recognized and neutralized by the targeting spacer, i.e. the CRISPR is more effective. This internal error tolerance makes it harder for phage to escape by mutating one bit of their protospacer for $l = 2$. If the number of mismatches is greater than one, the spacer provides no protection against the phage for $l = 2$. We show that there is little difference in the results for point mutation and recombination when $l = 1$. However, when $l = 2$, the difference in results between the point mutation and recombination becomes apparent. It is widely assumed that $l = 1$ describes phage recognition [8]. It seems likely, however, that a protospacer with a single mismatch would also be recognized, i.e. $l = 2$ should apply in at least some cases, and some evidence for single-mismatch recognition has been observed [9, 18].

The different CRISPR-evading strategies of recombination and mutation have minimal impact on the spacer diversity with respect to position, as shown by Fig.
2.8 in comparison to Fig. 2.5. For \( l = 2 \), the spacers are slightly more diverse when phage recombine than they when mutate. Although recombination allows phage to make a more diverse set of descendants than does point mutation, the observed effect in the diversity of CRISPR is small. Thus, spacer diversity is not a sensitive measure to distinguish different CRISPR-evading strategies.

At long times, the diversity of the leader-proximal spacers decreases. This is because the diversity of the phage population itself decreases for large time. This diversity profile is shown in Fig. 2.9.

We define “immunity” as a measure of the possibility that CRISPR will kill phage:

\[
\beta \sum_k \sum_{i,j} x_{i,j} v_k (\delta_{i,k} + \delta_{j,k})
\]

The higher the immunity is, the higher protection the spacer provides. Figure 2.11 shows that recombination gives phage more chance to survive and the CRISPR immunity is lower. When \( l = 1 \), immunity is similar whether pages escape by mutation or recombination, because the effectiveness of spacer is equal in regard to escape by point mutation or recombination. When \( l = 2 \), the immunity is higher against escape by mutation than it is against escape by recombination. Immunity decays more quickly with recombination rate than with mutation rate, Fig. 2.11.

2.3 Discussion

We have addressed whether or not bacteria population dependence should be included in the phage growth rate, \( r \). We have shown that a natural form of nonlinear growth dynamics makes no difference at long time in the regime where phage and bacteria coexist, although there is a slight difference at short time. Since we enforce coexistence, this detail is inessential under the conditions of our study. There are multiple strains of phage and bacteria, and most strains of the phage can grow in
Figure 2.8: Diversity at different positions of CRISPR at different times for $l = 2$ with a) mutation only and b) recombination only. In this case, CRISPR recognize phage with zero or one mismatch between the spacer in the bacterium and protospacer in the phage. The recombination rate per sequence per replication is $\nu = 0.01$, and $p_c = 0.5$. The other parameters are the same as in Fig. 2.4. Spacer diversity is not particularly sensitive to whether the phage evolve by mutation or recombination.

Figure 2.9: Diversity of the phage for $l = 2$ with mutation only. The parameters are the same as in Fig. 2.8.
Figure 2.10: Protection at different positions of CRISPR at different times for $l = 2$ with a) mutation only and b) recombination only. The parameters are as in Fig. 2.8.

nearly all strains of the bacteria in our simulations.

The diversity of the spacers at the leader-proximal end shown in Fig. 2.5 is higher than the diversity of the spacer at the leader-distal end. This result is consistent with experimental observations on different bacteria [7, 8, 44, 23, 42]. The difference in diversity between these two ends decreases as time elapses as the spacers fill in the CRISPR and the phage strains randomize due to mutation. This result shows the diversity of the spacers increases as the diversity of phage increases. This result is also observed in a more complex microbial community experimentally [45].

It has often been assumed that when the CRISPR is “full” and spacers are to be deleted, the oldest spacer is deleted, or the oldest spacer is more likely to be deleted. Not all mechanisms for spacer deletion are capable of such a biased removal [46]. An equal deletion probability for all spacers is a simpler and perhaps more biologically motivated assumption. We have shown that such a uniform deletion probability does give a spacer diversity which decreases with distance from leader sequence, in accord with observation. A uniform deletion rate may be a simple, yet representative model for spacer deletion.
We have quantified the impact of mutation and recombination on phage escape from CRISPR recognition. Mutation and recombination both allow phage to escape. So far, most theories have assumed that phages evolve only by point mutation. Here we have examined the effects of recombination on the coevolution process, complimenting previous theoretical studies [47, 48, 49, 50, 51, 52]. Data suggest that recombination is a significant driver of evolution [23]. To quantify the effectiveness of mutation versus recombination in phage escape, we defined a new quantity, “immunity,” the rate at which bacteria kill phages. This immunity is a good measure of the effectiveness of phage escape. By computing immunity, we quantified and compared the relative efficiencies of mutation and recombination for phage escape. There may be selective constraints on what mutations can occur in the viral protospacer. Consequently, phage need to find “viable” mutations. Recombination in the phages can combine beneficial or viable mutations. Furthermore, one mutation may not necessarily be enough to escape the host CRISPR immunity system, and it is possible that
greater than one mutations is needed in order for a phage to escape. For both of these reasons, recombination allows phage to escape CRISPR recognition more effectively than does mutation alone.

Differing immune pressures become distinguishable in the diversity measurements at long times. At short times, the diversity results for $l = 1$ in Figure 2.5 and for $l = 2$ in Figure 2.8 are similar. The results differ at longer times, $t \geq 1200$, in these two figures.

Interestingly, the leader-proximal spacers are less diverse in Fig. 2.8 for $l = 2$ than they are for $l = 1$. A lower diversity of these spacers is also observed for smaller mutation or recombination rates. When the phage is less able to escape the CRISPR, the diversity of the phage population decreases at long times. For this reason, the diversity of the spacers incorporated at later times, i.e. the leader-proximal spacers, is lower than that of spacers incorporated earlier, i.e. the spacers a bit farther from the leader.

If the bacteria are killed less by the phage, for example by having a more effective immune system, they are able to add a greater number of spacers and to fill up their CRISPR array more quickly. As the spacers fill in the leader-distal CRISPR positions, the diversity rises above the initial value of zero. It is for this reason that leader-distal diversity as a function of position for smaller mutation rate, smaller recombination rate, or larger $l$ are above those for higher mutation rate, higher recombination rate, or smaller $l$. The interplay between the decrease of phage population diversity at long times and the filling in of the CRISPR array leads to the non-monotonic diversity of spacers with position in Fig. 2.8. The protection as a function of position can also be non-monotonic, as is Fig. 2.10, due to a decreasing diversity of phage with time and the diversity of leader-distal spacers being greater than that of intermediate spacers.
Protection of CRISPR is a better measure to differentiate the two CRISPR-evading strategies of mutation or recombination. From the figures of protection versus position, we can see that when $l = 2$, the protection of CRISPR is lower when the phage recombine, Fig. 2.10b, than mutate, Fig. 2.10a. That is, recombination allows the phage to escape the CRISPR system more easily. This result illustrates that recombination is a more efficient CRISPR evading strategy for phage.

2.4 Conclusion

The CRISPR/Cas system plays a crucial role in bacteria and phage coevolution. By adding and deleting spacers, bacteria are evolving dynamically under the selection pressure imposed by phage undergoing point mutation and recombination. The stochastic model used in this work captures the essential features of the CRISPR/Cas system, giving rise to the fascinating characteristics coexisting bacteria and phage system. The rich variety of spacers within the CRISPR locus captures the history of bacteria and phage coevolution.

As the “ancient” winner with better fitness, the leader distal spacers are more homogeneous than the leader proximal spacers. This result has previously been observed under a wide range of model parameters [26]. Bacteria with more effective immune systems, or bacteria attacked by phage that mutate more slowly, have higher fitness and are able to more quickly fill their CRISPR array with spacers. This result is rather intuitive and expected to hold under rather general conditions[47, 48, 49, 50, 51].

Spacer diversity is not particularly sensitive to whether the phage evolve by mutation or recombination. This result may be a bit surprising. It is understood to be a result of recombination between two random phage strains almost always leading to a new phage strain, and, therefore, identical in effect to mutation.
Different mechanisms of spacer deletion subtly affect the distribution of spacers in CRISPR. Random deletion of spacers [7, 9] leads to a modestly slower rate of filling in the CRISPR array than does a mechanism of deleting only the leader-distal spacer. This result is because random deletion removes non-terminal spacers, which inhibits growth.

The protection or immunity that CRISPR confers upon bacteria is sensitive to the effectiveness of CRISPR-phage recognition, distinguishing between whether $l = 1$ and $l = 2$ mismatches are required for phage to escape recognition. Protection and immunity are also sensitive to the mechanism of phage escape, easily distinguishing different rates of phage evolution. Recombination is seen [23] to be more effective in allowing phage to escape CRISPR recognition when greater numbers of mutations are required for escape, $l = 2$. This result is simply because recombination is most often with a quite different strain, and so the produced recombinant has more contained variation than mutation would provide. It is likely that phage recombination is a significant generator of phage diversity in the wild.

2.5 Table of Parameters

The parameters shown in the above chapter and their meanings are in the following table.
Table 2.1: Table of parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{i,j}$</td>
<td>The population of bacteria containing spacers $i$ and $j$</td>
</tr>
<tr>
<td>$x_{\text{max}}$</td>
<td>The maximum bacteria population</td>
</tr>
<tr>
<td>$v_k$</td>
<td>The phage population containing protospacer $k$</td>
</tr>
<tr>
<td>$v_{\text{max}}$</td>
<td>The maximum phage population</td>
</tr>
<tr>
<td>$c$</td>
<td>The bacteria growth rate</td>
</tr>
<tr>
<td>$r$</td>
<td>The phage growth rate</td>
</tr>
<tr>
<td>$\beta$</td>
<td>The bacteria exposure rate</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>The probability of acquiring a new spacer</td>
</tr>
<tr>
<td>$\mu$</td>
<td>The mutation rate</td>
</tr>
<tr>
<td>$\nu$</td>
<td>The recombination rate</td>
</tr>
</tbody>
</table>
Chapter 3

Waves of Spacer Replacement in CRISPRs Acquired from Phages with Multiple Proto-spacers

In this chapter, we extend the model shown in Chapter 2 to two dimensions and add some new features, such as the multiple proto-spacer, to explore the spatial effect on the evolutionary dynamics of CRISPR and phages. Comparing with the results in Chapter 2, here we show the spatial distribution and correlation of bacteria and phages. We also explicitly show that CRISPR accelerates the co-evolution of bacteria and phages by comparing with the neutral model.

We introduce the multiple proto-spacer, two dimensional stochastic model of bacteria and phage coevolution in section 3.1. In section 3.2, we study the evolution of diversity in the spacers of the bacteria. We report how the spatial effect and multiple proto-spacer loci affect the evolutionary dynamics. In section 3.3, we compare our observations with experimental data. We discuss the relation of the CRISPR immunity with the number of proto-spacer loci in phages and the number of spacers in CRISPR. We conclude in section 3.4.

3.1 Methods

The model used here is based upon the method introduced by He and Deem [26], with several new features. First, we extended the model into two dimensions, where bacteria interact with nearby phages. Second, the growth of phages was considered in more detail, including the phage latent time and burst size. Additionally, we here
consider multiple proto-spacers.

The system of bacteria and phage was simulated in a two-dimensional \( L \times L \) grid. The grid size was such that each site contains either zero or one bacterium. The maximum number of phages at each grid site was 20. Phages could have up to 30 proto-spacer loci. The length of each spacer and proto-spacer was 10 base pairs. This model is continuous in time. The Lebowitz-Gillespie algorithm was used to exactly simulate the stochastic events. There is no finite \( \Delta t \) error in this approach. Averages computed by the Lebowitz-Gillespie are exact, in the limit of an infinite number of simulation runs. [53, 54]. A finite number of simulations gives an approximation to the average, with an estimate of the standard error. Our figures contain error bars indicating the standard estimate of the error. The boundary conditions were periodic.

The bacterium and phage on the same site could interact. Each bacterium had two states, infected or healthy. The healthy bacteria could be infected by phages. Healthy bacteria could reproduce if there was an empty neighboring grid. Daughter cells were randomly placed into an available neighboring site. If all of the neighbor sites were occupied, this healthy bacterium could not reproduce. The infected bacteria lysed after the latent time. After lysis, phages spread by diffusion or disappeared by decay. Spacers were categorized by spacer type. Identical spacers were grouped into one spacer type. We kept track of the frequency of each spacer type, and we calculated the turnover and the diversity of spacers.

3.1.1 Events

These events described above occur simultaneously, each at a specified rate. The Lebowitz-Gillespie algorithm was used to compute averages. The events are grouped into seven categories:
1. **Bacteria reproduction**: A healthy bacterium could reproduce if there was an empty neighboring grid site. The rate of one healthy bacterium reproducing was $A$. After the bacterium reproduced, one empty neighbor site was randomly chosen, and the daughter cell was placed on it.

2. **Bacteria infection**: A healthy bacterium could be infected by phages at the same site. The adsorption rate of phages to a healthy bacteria was $\beta \times v$, where $\beta$ is the exposure rate and $v$ is the number of free phages on this site. Upon phage attack, the bacteria had a probability $\gamma$ to acquire a new spacer from the invading phage genome. If the phage had multiple proto-spacer loci, say $p$ proto-spacers, every proto-spacer has the probability $\gamma/p$ to be acquired. The newly acquired spacer was always inserted at the leader-proximal end of CRISPR. The alternative event, with probability $1 - \gamma$, was no incorporation of a proto-spacer. If the CRISPR system of the bacteria recognized the phage, i.e., any spacer in the CRISPR matched any proto-spacer of the phage, the bacterium was still healthy and the phage was killed by the bacterium. Otherwise, this bacterium became infected. Fig. 3.1 shows each process during the bacteria infection event.

3. **Bacteria lysis**: Infected bacteria had a rate $1/\tau$ to lyse, where $\tau$ is the latent time. When the infected bacteria lysed, $b$ new phages came out. The number of the new phages, $b$, is called the burst size. The new phages spread on the nearest eight sites and the site where they came out, which is shown in Fig. 3.2. The number of phages on each site increased by $b/9$. If the number of phages exceeded the maximum phage population on each site, the new phages replaced some randomly chosen old phages. Each of the newborn phages could have point
Figure 3.1: Bacteria infection events.

Figure 3.2: The phages spread out after lysis.
mutations or recombination upon bacterial lysis. The recombination procedure in our model is as follows: when a new phage was born, it had a probability \( \nu \) to swap its genetic materials with a random phage in the current population. When two phages recombined, first we randomly chose a sequence to start copying. The copy process had a probability \( 1 - p_c \) to continue on that sequence or probability \( p_c \) to jump to the other sequence, until a complete recombined sequence was created. The recombined sequence was assigned to the new phage.

4. **Phage diffusion**: Phages could jump to neighboring sites. The diffusion constant is \( \alpha \). The rate of one phage jumping to a neighboring site is \( \alpha/l^2 \), where \( l \) is the grid spacing (see the Supporting Information). If the population of phages in the neighboring site was already the maximum value, one random phage in this neighboring site was replaced.

5. **Phage degradation**: Each phage had a decay rate \( \delta \). If a phage decayed, the number of phages on this site decreased by one.

6. **Phages killed by CRISPR system**: When one phage injected its gene into a bacterium, this phage was degraded by the CRISPR system if one spacer in the CRISPR of the bacterium matched any proto-spacer in the genome of the phage.

7. **CRISPR incorporates new spacers**: Bacteria had a probability \( \gamma \) to acquire a new spacer into the spacer array after injection of the phage gene. The newly incorporated spacer was inserted at the first position of the spacer array. Old spacers were shifted to the distal end. The number of spacers in CRISPR was limited to one to match the experimental conditions of [41]. We also examined the case of four spacers. The oldest spacer was deleted when bacteria acquired a
new spacer into the CRISPR and the maximum number of spacers was already
in the CRISPR.

3.1.2 Parameters and initial conditions

Published data were used to determine the values for most of the parameters, namely
the grid spacing, the growth rate of bacteria, the phage adsorption rate, the latent
time of phages, the phages burst size, the diffusion rate of phage, number of proto-
spacers in phage, and number of spacers in CRISPR. For the parameters without
published values, i.e. the rate of acquiring new spacers and the phage decay rate,
biologically reasonable values were estimated.

All the bacteria initially had no spacers, and some infected bacteria were initially
placed on the grids. Several initial geometric conditions for the infected bacteria were
examined: infected bacteria randomly placed on the lattice initially, infected bacteria
placed on several spots, and infected bacteria placed on one or two strips initially,
which are shown in Fig. 3.3.

The phage and bacteria became well mixed at long time due to diffusion. There
was almost no difference between the steady state distribution arising from these dif-
ferent initial geometric conditions. The initial geometric condition shown in Fig. 3.3c
is used for most of the results presented here.

3.2 Results

3.2.1 Daily waves of replacement in the CRISPR spacers

Due to the co-evolution of the bacteria and phage, the distribution of spacers in
CRISPR changes rapidly. Fig. 3.4a shows the daily change of the 10 most abundant
spacers. All the bacteria had no spacers initially. The bacteria that acquired one spacer had a higher fitness. They grew faster than the bacteria whose CRISPRs were empty. Therefore, a few spacers dominated the distribution of spacers in the first day. Those bacteria that acquired spacers imposed a selection pressure on the phages. This selection changed the distribution of proto-spacers. The dominant bacteria of the first day gradually lost their advantage and were replaced by bacteria incorporating other spacers at subsequent days. This dynamics lead to successive waves of dominant spacers in the CRISPR population.

3.2.2 Spacer turnover

Spacer turnover is defined as a measure of the daily replacement of the spacer distribution [41]. It is one minus the sum of the minimum percentage of each spacer at consecutive time points. Defining the probability of spacer $i$ at time $t$ by $p_i(t)$, the spacer turnover is $1 - \sum \min[p_i(t), p_i(t + 1)]$. Using this definition, the value of turnover is bounded between zero and one. If the spacer distribution did not change during the two days, the turnover was zero. A higher value of the turnover means the distribution of spacer changes more drastically. Fig. 3.4b shows the spacer turnover during 10 days. These results are similar to the data that motivated this theory [41].
Figure 3.4: (a) **Daily abundance of the most 10 abundant spacers.** Here the number of spacers in CRISPR is one, number of proto-spacers in phage genome is 30, length of every proto-spacer is 10 bp, bacterial growth rate is 0.04 min\(^{-1}\), mutation rate of phage is 0.002 per generation per proto-spacer, and probability of incorporating a new spacer is 0.001. The values of other parameters are shown in Table. 3.1. The results are similar to the experimental data of [41]. (b) and (C) **Spacer turnover during the first 10 days predicted by simulation and measured in experiment.** (b) The simulation result of spacer turnover during the first 10 days. Parameters are the same as those in (a). (c) The experiment data of spacer turnover [41]. (d) and (e) **Daily abundance of the most 10 abundant spacers and spacer turnover during the first 10 days when there is one proto-spacer locus.** Other parameters are the same as those in (a).

### 3.2.3 Sensitivity to parameter values

The effect of changing several parameters was investigated. The number of spacers in CRISPR, the number of proto-spacers in the phage genome, the mutation rate
of phages, and the mutation mechanism were all changed. The influence of these parameters is shown below.

**Number of spacers**

The number of spacers in CRISPR was increased from the experimental value of one [41] to a value of four. Fig. 3.5 shows the spacer distribution when the number of spacers is four. Compared to Fig. 3.4a, the diversity of spacers is higher at long time. This is because when the number of spacers is four, the immune pressure of CRISPR on the phages is greater, and the phages evolve more.

![Graph showing daily abundance of the 10-most-abundant spacers](image)

Figure 3.5: **Daily abundance of the 10-most-abundant spacers.** The number of spacers is four. Other parameters are the same as those in Fig. 3.4.

**Number of proto-spacer loci**

The number of proto-spacer loci in the phage genome was varied between one and 30. Fig. 3.4d shows the waves of replacement in CRISPR spacers, and Fig. 3.4e shows the turnover of spacers, when phages had one proto-spacer locus. Comparing
Fig. 3.4de with Fig. 3.4ab, the replacement waves and turnover are different during the first four days, and the spacer population structure is different at later days. When phages had one proto-spacer locus, there was only one dominant spacer in the bacteria population, and it decreased slowly in the first several days.

Fig. 3.6 shows how the phage population changes when the number of proto-spacer loci is varied. Increasing the number of proto-spacers made it harder for the phage to escape recognition by CRISPR, and the steady state number of phage was smaller.

*Figure 3.6 : The population of phages for one proto-spacer locus and 30 proto-spacer loci. The number of spacers is four, the probability of acquiring a new spacer is 0.01, and other parameters are the same as those in Fig. 3.4. Error bars are one standard error.*

**Mutation rate**

Increasing the mutation rate of the phages made it easier for the phages to escape from CRISPR recognition. Fig. 3.7 shows the population of phages for different mutation rates. A higher mutation rate led to a greater number of phages.
Figure 3.7: The population of phages for different mutation rates of phages. Phages are placed on a strip initially. The population of phages in the area near the strip was tracked. The mutation rates are from 0.0002 to 0.1 per generation per proto-spacer, number of spacers is one, number of proto-spacer is one, probability of acquiring a new spacer is 0.1, and other parameter are the same as those in Fig. 3.4. Error bars are one standard error.

Spatial correlation

Because the model was in two dimensions, bacteria formed clusters, and phages gathered around the cells if the phage diffusion rate was small. Fig. 3.8ab shows one snapshot of the evolution system. To quantitatively study this distribution, the spatial correlation of healthy and infected bacteria was calculated. The correlation is defined by

\[
C_{\alpha,\beta}(i, j) = \sum_{(m,n)} \frac{x_{\alpha}(m, n)x_{\beta}(m + i, n + j)}{\sqrt{n_\alpha} \sqrt{n_\beta}},
\]

(3.1)

where \(x_{\alpha}(m, n)\) and \(x_{\beta}(m, n)\) are the number of healthy bacteria \((x_1)\) or the number of infected bacteria \((x_2)\) at site \((m, n)\), and \(n_\alpha\) and \(n_\beta\) are the total number of healthy bacteria \((n_1)\) or the total number of infected bacteria \((n_2)\). Fig. 3.8c-e shows these correlations.
Figure 3.8: (a) The distribution of phages, (b) the distribution of bacteria. Green is healthy bacteria, red is infected bacteria, and blue is an empty site. The correlation of (c) healthy with healthy bacteria, (d) infected with infected bacteria, and (e) healthy with infected bacteria. The number of spacers is one, the number of proto-spacer is 30, bacterial growth rate is 0.01 min$^{-1}$, phage diffusion rate is 0.32 $\mu$m$^2$·min$^{-1}$, the probability of acquiring a new spacer is 0.001, mutation rate of phages is 0.001 per generation per proto-spacer, and other parameters are the same as those in Fig. 3.4.

From the correlation matrix $C_{\alpha,\beta}(i,j)$, the correlation was calculated as a function of radius, $C_{\alpha,\beta}(r)$. Fig. 3.9 shows the function $C_{1,2}(r)$. The function of $C_{1,2}(r)$ depends on the parameters of the co-evolving system, such as the exposure rate. Fig. 3.10 shows $C_{1,2}$ for different exposure rates. For a higher exposure rate, there were more infected bacteria around the healthy bacteria, and a higher peak in $C_{1,2}$ was observed.
Figure 3.9: The correlation of healthy and infected bacteria as a function of radius. Parameters are the same as those in Fig. 3.4.

Figure 3.10: The correlation of healthy and infected bacteria as a function of radius for different infection rates. The exposure rates are $\beta = 0.0008 - 0.001\text{min}^{-1}$. Other parameters are the same as those in Fig. 3.4.

**Recombination vs. mutation**

It has been shown that if mismatch is allowed during the CRISPR interference stage, recombination is a more efficient CRISPR-evading strategy than is point mutation.
because recombination generates more diverse offspring than does point mutation, when parents are diverse [27]. Here we saw that if the number of proto-spacer loci was greater than one, even when no mismatch is allowed, recombination was still more efficient than point mutation at facilitating phages to escape recognition by CRISPR. Fig. 3.11 shows the phage population when phages had only recombination or only point mutation. The population was higher when phages only had recombination.

Figure 3.11: The population of phages when phages only had recombination or only had point mutation. Phages were placed in a strip initially. The population of phages near the strip was tracked. The number of spacers is four, number of proto-spacers is two, probability of acquiring a new spacer is 0.1, recombination rate and mutation rate are 0.005 per proto-spacer per generation, and other parameters are the same as those in Fig. 3.4. Error bars are one standard error.

To measure the efficiency of recombination and point mutation, the fraction of phages on sites with bacteria that do not have any proto-spacers recognized by any of the bacteria on each site, $S$, was defined as follows:

$$S = \frac{\sum_{i,j} \sum_l (v_{i,j}^{l}(\tilde{k}^l)x_{i,j}(\tilde{m}) \prod_{\alpha,\beta} (1 - \delta_{\alpha,\beta}))}{\sum_{i,j} \sum_l v_{i,j}^{l}(\tilde{k}^l)x_{i,j}(\tilde{m})},$$

(3.2)
where $v_{l,j}(\vec{k})$ is the $l^{th}$ phage in site $(i, j)$ with proto-spacer array $\vec{k}$, $x_{i,j}(\vec{m})$ is the bacterium on site $(i, j)$ with spacer array $\vec{m}$, $\delta_{k_{\alpha},m_{\beta}}$ is the Kronecker delta function, $k_{\alpha}^l$ is the $\alpha^{th}$ proto-spacer in $\vec{k}$, and $m_{\beta}$ is the $\beta^{th}$ spacer in $\vec{m}$. A higher value of $S$ means the bacteria were more susceptible to phages. For example, if all spacers in CRISPR were empty, $S$ would be one, meaning that all phages could infect the bacteria. When bacteria acquired spacers that could protect from most of the phages, this quantity would approach zero because almost no phages could infect bacteria.

How the fraction of phages that can infect bacteria depends on recombination and point mutation is shown in Fig. 3.12ab for one type of phage initially. Point mutation was more effective than recombination for phage escape, because there was no diversity to recombine initially. When there are two types of phage initially, recombination was more effective than point mutation for phage escape. The comparison is in Fig. 3.12cd. When there were more types of phage at first, say 20 types, recombination and point mutation had a similar escape efficiency, Fig. 3.12ef.

To show the different evolutionary dynamics arising from recombination or point mutation more clearly, we show how $\langle S_{\text{recombination}} \rangle - \langle S_{\text{point mutation}} \rangle$ depends the number of initial viral strains. From Fig. 3.13, recombination was less effective than point mutation only when there was one viral strain initially. When there were multiple viral strains, phages undergoing recombination could more easily escape CRISPR recognition. This advantage of recombination decreased as the number of initial viral strains increased.

### 3.2.4 Neutral model

When bacteria acquired a new spacer, the old spacers were shifted to the distal end of the spacer array. How does this shift mechanism interact with replication to affect
Figure 3.12: **Recombination versus point mutation for phage population and infection probability.** In (a) and (b), there was one type of phages initially. In (c) and (d), there were two types of phages initially. In (e) and (f), there were 20 types of phages initially. The number of spacers was four, number of proto-spacers was two, mutation rate and recombination rate are 0.0025 per proto-spacer per generation. Other parameters are the same those in Fig. 3.11. Error bars are one standard error.

the diversity of spacers? Can it lead to a decrease of diversity from the leader end to the distal end? Here we consider what happened when the bacteria incorporated a novel spacer from a “proto-spacer bank” and the bacteria grew freely without phage pressure. This is a neutral model. Even without selection, the diversity of spacers decreased from the leader end to the distal end, which is shown in Fig. 3.14a.

A master equation describes this neutral random process:
Figure 3.13: The difference in phages not recognized by CRISPR for recombination or point mutation dynamics, $S_{\text{recombination}} - S_{\text{point mutation}}$, versus the number of initial viral strains. Other parameters are the same as those in Fig. 3.12.

Figure 3.14: (a) A neutral model of the diversity of spacers. There are four spacers in CRISPR, $N = 10000$, $L = 10$, and the ratio of bacteria growth rate to the rate of acquisition of new spacers, $A/\gamma$, is 10. Proto-spacers are randomly acquired into the spacer array. (b) The comparison of the theoretical and simulation results for the neutral model. Here $N = 50000$, and $L = 7$.

\[
\frac{dP\{n_{a,b}\}}{dt} = A \sum_{(a,b)} (n_{a,b} - 1) \sum_{(\alpha,\beta) \neq (a,b)} \frac{n_{\alpha,\beta} + 1}{N} P(n_{a,b} - 1, n_{\alpha,\beta} + 1) 
- A \sum_{(a,b)} n_{a,b} \sum_{(\alpha,\beta) \neq (a,b)} \frac{n_{a,\beta}}{N} P(n_{a,b}, n_{a,\beta}) 
+ \frac{\gamma}{2} \sum_{(a,b)} \sum_{\alpha} (n_{b,\alpha} + 1) P(n_{a,b} - 1, n_{b,\alpha} + 1) 
- \frac{\gamma}{2} \sum_{(a,b)} \sum_{\alpha} n_{b,\alpha} P(n_{a,b}, n_{b,\alpha})
\]  

(3.3)
where $n_{a,b}$ is the number of bacteria with the first spacer $a$ and the second spacer $b$. $P\{n_{a,b}\}$ is the probability of the configuration $\{n_{a,b}\}$. Here $A$ is the growth rate of bacteria, $N$ is the total population of bacteria, $L$ is the length of spacer, and $\gamma$ is the rate of acquiring a new spacer. From a moment analysis of Eq. 3.3, one obtains

$$\frac{d\langle n_{\alpha\beta} \rangle}{dt} = \frac{\gamma}{2L} \sum_{a'} \langle n_{\beta a'} - n_{\alpha\beta} \rangle$$

Therefore, when the system reaches steady state, $\langle n_{\alpha\beta}(\infty) \rangle = \langle n_{\beta}(\infty) \rangle \frac{2}{N}$.

The diversity of the first spacer is

$$D(t) = -\sum_a \langle \frac{n_a(t)}{N} \ln \frac{n_a(t)}{N} \rangle$$

$$= -\sum_a \langle \frac{n_a(t) + \delta n_a(t)}{N} \ln \frac{n_a(t) + \delta n_a(t)}{N} \rangle$$

$$\approx -\sum_a \frac{\langle n_a(t) \rangle}{N} \ln \frac{\langle n_a(t) \rangle}{N} - \sum_a \frac{\langle (\delta n_a(t))^2 \rangle}{2N \langle n_a(t) \rangle}$$

$$= D_0 - \frac{1}{2N} \sum_a \frac{\langle (\delta n_a(t))^2 \rangle}{\langle n_a(t) \rangle}$$

$$= D_0 - \frac{1}{2N} \sum_a \frac{\langle n_a^2(t) \rangle - \langle n_a(t) \rangle^2}{\langle n_a(t) \rangle}$$

$$= D_0 - \frac{1}{2N} \sum_a \frac{\langle n_a^2(t) \rangle - N^2/4L}{N/2L}, \text{ when } t \rightarrow \infty$$

$$= D_0 - \left( \frac{4L}{2N^2} \langle n_a^2 \rangle - \frac{1}{2} \right), \text{ when } t \rightarrow \infty$$

So the decrement of the diversity due to finite $N$ is $\Delta D = \frac{4L}{2N^2} \langle (\delta n_a)^2 \rangle$. As shown in the Supporting Information, the result for the first spacer is $\Delta D = \frac{(2L-1)(A+\gamma)}{4(A+N\gamma)}$.

Similarly, the result for the second spacer of a two spacer model is

$$\Delta D_2 = \Delta D + \frac{(2L-1)(N-1)AN\gamma^2}{4(A+N\gamma)^3}$$

When $L > 0$ and $N > 1$, $\Delta D_2$ is always larger than $\Delta D$. Fig. 3.14b shows convergence of simulated values to the theoretical result for large $A/\gamma$. 
3.2.5 CRISPR accelerates the co-evolution

The CRISPR system imposes a selection pressure on the phage, and it increases the evolution of the phage and CRISPR. To see this clearly, the neutral model was used, and the CRISPR interference was inactivated, i.e., the CRISPR could acquire a new spacer from the phage but the CRISPR did not degrade the phages. The diversity of spacers increased far more slowly when CRISPR immunity was inactivated, Fig. 3.15. The diversity of phages also increased more slowly when CRISPR was inactivated, Fig. 3.16. Thus the CRISPR system changed the structure of the phage population and pushed the phages to evolve rapidly.

Figure 3.15: The diversity of spacers versus time when CRISPR was (a) active and (b) inactive. The number of spacers is four, number of proto-spacers is one, mutation rate of phages is 0.02 per proto-spacer per generation, probability of incorporating a new spacers is 0.1, and other parameters are the same as those in Fig. 3.4. Error bars are one standard error.

3.3 Discussion

The turnover in Fig. 3.4e, with a single proto-spacer in contrast to the experimental conditions, does not agree with the experimental data shown in Fig. 3.4c. First, under the conditions of the experiment in which there is only one strain of phage, the turnover for a single proto-spacer model starts near zero but is non-zero for
Figure 3.16: The diversity of phages versus time when CRISPR was active and inactive. The number of spacers is four, number of proto-spacers is one, mutation rate of phages is 0.02 per proto-spacer per generation, probability of incorporating a new spacers is 0.1, and other parameters are the same as those in Fig. 3.4. Error bars are one standard error.

multiple proto-spacers. This is a qualitatively different result, with the latter in agreement with experiment. Second, the steady state turnover is lower in the single proto-spacer versus the multiple proto-spacer model. This difference is perhaps more quantitative, with the latter in agreement with experiment. Note that the first several days are crucial to the coevolution of bacteria and phages because bacteria establish the immunity against phages during the first several days. Different spacer contents in the first 3 or 4 days implies a different bacterial population structure at later days. For these reasons, we conclude that multiple proto-spacer loci are needed to accurately describe this experimental system.

Fig. 3.11 shows that recombination is a more effective strategy for phage escape from CRISPR recognition that is point mutation if there are multiple proto-spacers. This is because recombination can change several proto-spacers at one time, increasing
the probability for phage to escape recognition by CRISPR. The performance of recombination versus point mutation depended on initial conditions. If there was only one type of phage initially, since recombination cannot generate new genotypes, point mutation was more efficient than recombination. Indeed, if no new genotypes were generated, no matter how large the recombination rate was, phages would go to extinction eventually.

From Fig. 3.14a, the spacer diversity decreased from the leader end to the distal end even in a neutral model. This was a finite population effect, however, and this effect was quite small. When the bacterial population became very large, the decrement of the diversity was almost zero, Fig. 3.18. Note from Eq. 3.6, when $N \to \infty$, $\Delta D = O\left(\frac{1}{N}\right)$. Therefore, the decrease of the spacer diversity in Figs. 3.4–3.12 was mainly due to the coevolution of bacteria and phage [27], not because of the stochastic neutral mechanism in the acquisition of spacers.

Several models have been established to study the impact of CRISPR/Cas system on the coevolution of bacteria and phages. Haerter et al. first took spatial effects into account in their model [32, 33]. In their model, however, phages have only one proto-spacer. Thus, their model cannot capture the dynamics of spacer turnover shown in Fig. 3.4b. Childs et al. introduced an ecological and mean field model to analyze how CRISPR immunity to maintain the diversity of bacteria and phages [28]. However, this model did not include recombination of phage, and so it could not elucidate differences in dynamics induced by recombination and point mutation of proto-spacers, Fig. 3.11–3.13. Weinberger et al. used a detailed population-genetic model to study why leader-distal spacers are more conserved, arguing that selection plays a major role [34]. We here calculate the effect of spacer coalescence explicitly, showing decreasing of spacer diversity in a neutral model and comparing the effects
of selection and genetic drift rigorously.

3.3.1 Spacer distribution for different phage diffusion rates

Since this is a two dimensional model, the coevolution of bacteria and phages is expected to be affected by spatial effects. The distribution of spacers for different diffusion rates was investigated. Fig. 3.19 shows the comparison. If the phage diffusion rate was relatively large, the system was well mixed. When the mutant phages emerged, they diffused to the whole grid rapidly, and bacteria had a higher chance to interact with different types of phages. The bacteria with the most abundant spacer during the first day faced a larger selection pressure of the phages. Therefore, the highest value of the percentage of the most abundant spacer during the first day was lower than that when phage diffusion rate was smaller. As time elapsed, bacteria and phages also tended to mix even when the diffusion rate was small. So a steep drop is observed in Fig. 3.19b. At long times, spacer distributions were similar in Fig. 3.19ab because bacteria and phages were well mixed in both cases. This figure shows that spatial effects do influence the dynamics of the system. A lower diffusion rate makes the mixing of phages and bacteria slower. The increased clustering of phages and bacteria due to the lower diffusion rate of phages causes the system to evolve more slowly and causes a larger turnover of a limited number of dominant phages to occur at early times.

3.3.2 Spacer turnover

Due to the mutation of phages and acquisition of new spacers by the bacteria, the system coevolved rapidly. This coevolution is evident in Fig. 3.4a and 3.4b. Waves of replacement in the distribution of dominant spacers were seen. These results,
Fig. 3.4b, are similar to experimental data, Fig. 3.4c [41]. The spacer turnover as a function of time was similar for a wide range of parameter values. Fig. 3.20–3.23 shows similar spacer turnovers for different numbers of proto-spacers, probabilities of acquiring new spacers, and mutation rates of proto-spacers.

These waves of diversity replacement observed in the spacer abundance and quantified by spacer turnover were due to the interaction of spacers in CRISPR and proto-spacers in phage genome, \textit{i.e.}, the competition between the immune system of bacteria and the mutation and recombination of phages. At long times, the system tended to evolve to a uniform steady state, due to the mutation of phages. This is because almost every viral strain has the same fitness, and therefore the turnover of spacers reaches a constant steady state value at long times. The bacteria that gained spacers had a larger fitness. They grew faster than those bacteria whose CRISPR were empty, and they eventually dominated the bacteria population. At the same time, the fitness of phages decreased as bacteria incorporated more spacers. The population of phages, therefore, decreased. If phages did not mutate or the mutation rate was quite small, the phage could go extinct. When the mutation rate was above a threshold, the phages recovered.

\subsection*{3.3.3 Immunity}

The immunity of CRISPR is related to the number of spacers in CRISPR and the number of proto-spacers in phages genome. If the number of spacers in CRISPR increased, bacteria had a longer memory of phage challenges. Therefore, the bacteria were more immune to the phages. On the other hand, if the number of proto-spacer loci in the phage genome increased, the phages were more vulnerable to the CRISPR system because they had a higher chance to be recognized by bacterial immune sys-
tem. These immune pressures were reflected in the population of bacteria and phages. A higher population of bacteria and lower population of phages implies the immunity was stronger, and vice versa. From Fig. 3.6 one can observe that increasing the number of proto-spacers made the phage more susceptible to the CRISPR immune system.

### 3.3.4 Peak in phage population

There is a peak in the phage population growth curve in Fig. 3.7. The phages experience exponential growth and almost reach the maximum population size rapidly. This is because the CRISPR in all bacteria were empty at first, and so all the bacteria were initially susceptible to phages. As bacteria incorporated new spacers, the surviving bacteria gained immunity to the phages. With a higher fitness, these surviving bacteria reproduced. At the same time, the phage population could decrease, depending on the mutation rate and the probability of incorporating new spacers. These competing effects could generate a peak in phage population. The shape of the peak mainly depended on the probability of acquiring new spacers and the mutation rate of phages. For a higher probability of acquiring new spacers, the decrease of phage population was more drastically, Fig. 3.24. For a higher mutation rate of phages, the decrease of phage population was slower, Fig. 3.7.

### 3.4 Conclusion

We have shown that when literature values of the parameters are used, experimental data [41] on spacer abundance and turnover can be explained by an explicit, spatially-dependent model of bacteria and phage. The coevolution in this system is driven by CRISPR immune pressure on the multiple proto-spacer loci in the phages and by
the phage infection pressure on the CRISPR. We showed that the evolution rate and diversity of the phages in this system are significantly larger than that in a neutral model. We also showed that recombination of different proto-spacer loci leads to recombination being a more effective escape strategy than is point mutation for phage to escape from CRISPR pressure when there is an intermediate amount of genetic diversity in the phage population. Additionally, we explicitly calculated the diversity of spacers in a neutral model. This calculation showed that selection, rather than genetic drift, was the main reason for diversity of spacers to decrease from the leader end to the distal end in all but the smallest phage populations.

We make a number of predictions about the population dynamics of the phage and bacteria system. The phages and bacteria are in correlated patches in space, with phages around islands of healthy bacteria. We predict that as the number of spacers in CRISPR increases, the immune pressure on the phages increases and so does the phage diversity. Similarly, increasing the number of proto-spacer loci increases the immune pressure on the phages, and the number of phages decreases. Conversely, increasing the phage mutation rate lead to a higher number of phages. Recombination was an even more efficient strategy for further evolution of a diverse population of phages. Bacteria with no initial immunity to a new environment of phages were depressed in population numbers, but could recover if CRISPR was sufficiently adaptive. In this case, there was typically a dip in the phage population as the CRISPR pressure built, before the phage began to diversify and a steady state population size evolved. CRISPR dramatically affects the phage rate of evolution, significantly increasing it beyond the neutral rate.
Table 3.1: Table of parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>Grid size</td>
<td>40 l</td>
</tr>
<tr>
<td>$l$</td>
<td>Grid spacing</td>
<td>4 μm</td>
</tr>
<tr>
<td>$A$</td>
<td>Bacterial reproduction rate</td>
<td>0.01–0.04 min$^{-1}$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Bacterial exposure rate</td>
<td>0.001 min$^{-1}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Probability of acquiring new spacers</td>
<td>0.001–0.1</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Latent time</td>
<td>100 min</td>
</tr>
<tr>
<td>$b$</td>
<td>Phage burst size</td>
<td>90</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Phage diffusion constant</td>
<td>3.2 μm$^2 \cdot$ min$^{-1}$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Phage decay rate</td>
<td>0.01 min$^{-1}$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Mutation rate</td>
<td>0.0001 – 0.1 proto-spacer$^{-1}$ generation$^{-1}$</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Recombination rate</td>
<td>0.0001 – 0.1 proto-spacer$^{-1}$ generation$^{-1}$</td>
</tr>
<tr>
<td>$p$</td>
<td>Number of proto-spacers in phage</td>
<td>1–30</td>
</tr>
<tr>
<td>$q$</td>
<td>Number of spacers in CRISPR</td>
<td>1–4</td>
</tr>
</tbody>
</table>

The grid spacing is roughly the size of a bacterium; the value of $A$ is estimated from [55]; $\beta$, $\tau$, and $b$ are estimated from [56]; $p$ and $q$ are from [41]; $\mu$ and $\nu$ are from [34, 36]; the calculation of $\alpha$ is given in the Supporting Information. Biologically reasonable values of $\gamma$ and $\delta$ are used here.

3.5 Table of Parameters

3.6 Supporting Information

3.6.1 Neutral model

For the first spacer, one can derive from Eq. 3.3 an equivalent, simpler master equation for the neutral model,

$$
\frac{dP(n_a)}{dt} = A \sum_a (n_a - 1) \sum_{b \neq a} \frac{(n_b + 1)}{N} P(n_a - 1, n_b + 1) - A \sum_a n_a \sum_{b \neq a} \frac{n_b}{N} P(n_a, n_b) \\
+ \frac{\nu}{2T} \sum_a \sum_{b \neq a} (n_b + 1) P(n_a - 1, n_b + 1) - \frac{\nu}{2T} \sum_a \sum_{b \neq a} n_b P(n_a, n_b)
$$

(3.7)
From Eq. 3.7, one can obtain \( \langle n_{a'} \rangle \) and \( \langle n_{a'}^2 \rangle \),

\[
\frac{d\langle n_{a'} \rangle}{dt} = \frac{\nu}{2L} \sum_{\{n_a\}} \sum_{a \neq a'} (n_a - n_{a'}) P(n_a, n_{a'})
\tag{3.8}
\]

\[
\frac{d\langle n_{a'}^2 \rangle}{dt} = \frac{2A}{N} \sum_{\{n_a\}} n_{a'} \sum_{a \neq a'} n_a P(n_a, n_{a'})
\]

\[
+ \frac{\nu}{2L} \sum_{\{n_a\}} (-2n_{a'} + 1) n_{a'} \sum_{a \neq a'} P(n_a, n_{a'})
\]

\[
+ \frac{\nu}{2L} \sum_{\{n_a\}} (2n_{a'} + 1) \sum_{a \neq a'} n_a P(n_a, n_{a'})
\tag{3.9}
\]

Simplifying Eq. 3.8 and Eq. 3.9, one obtains

\[
\frac{d\langle n_{a'} \rangle}{dt} = \frac{\nu}{2L} \sum_a (\langle n_a \rangle - \langle n_{a'} \rangle)
\tag{3.10}
\]

\[
\frac{d\langle n_{a'}^2 \rangle}{dt} = \frac{2A}{N} \sum_{a \neq a'} \langle n_a n_{a'} \rangle + \frac{\nu(2L - 1)}{2L} \langle -2n_{a'}^2 + n_{a'} \rangle
\]

\[
+ \frac{\nu}{2L} \sum_{a \neq a'} \langle 2n_a n_{a'} + n_a \rangle
\tag{3.11}
\]

The solutions of Eq. 3.10 and Eq. 3.11 are

\[
\langle n_a \rangle = \frac{N}{2L}, \quad \langle n_{a'}^2 \rangle = \frac{N^2(2L^2 + 2L - 1) + \nu N}{4L (A + \nu N)}
\]

when \( t \to \infty \).

For the second spacer, from a moment analysis of the master equation (Eq. 2), one obtains

\[
\frac{d\langle n_{a\beta n_{\gamma\delta}} \rangle}{dt} = -\frac{2A}{N} \langle n_{a\beta n_{\gamma\delta}} \rangle - 2\nu \langle n_{a\beta n_{\gamma\delta}} \rangle
\]

\[
+ \frac{\nu}{2L} \sum_{a'} \langle n_{a\beta n_{\delta a'}} \rangle + \frac{\nu}{2L} \sum_{a'} \langle n_{\gamma\delta n_{b\alpha'}} \rangle
\]

\[- \delta_{a\beta} \langle n_{a\beta} \rangle - \delta_{b\gamma} \langle n_{\gamma\delta} \rangle \]

\[+ \delta_{a\gamma} \delta_{\beta\delta} \left( 2A + \frac{\nu}{2L} \sum_{a'} \langle n_{b\alpha'} \rangle + \nu \langle n_{a\beta} \rangle + \delta_{a\beta} \frac{2\nu}{2L} \langle n_{a\beta} \rangle \right) \]

\tag{3.12}
Solving Eq. 3.12, one obtains
\[
\langle n_{\alpha\alpha}^2 \rangle = \frac{(A + \frac{2L-1}{2}\nu)N^2}{(A + N\nu)4L} + \frac{N\nu}{2L} \left[ \frac{2^{L+1}AN^2 + \nu N^2(2L - 1)}{2^{3L+1}(A + \nu N)^2} \right.
\]
\[
+ \left. \frac{\nu N^3((2L + 1)A + (2L - 1)\nu + 2\nu N)}{2^{3L+1}(A + \nu N)^3} \right]
\]
\[
\langle n_{\alpha\beta}^2 \rangle = \frac{(A + \nu)N^2}{(A + N\nu)4L} + \frac{N\nu}{2L} \left( \frac{\nu N^3 - \nu N^2(2L + 1)}{2^{3L+1}(A + \nu N)^2} + \frac{\nu N^3(2L A + 2L\nu - \nu + \nu N)}{2^{3L+1}(A + \nu N)^3} \right)
\]
\[
\langle n_{\alpha\alpha}n_{\beta\alpha} \rangle = \frac{N\nu}{2L} \left( \frac{\nu N^3 + 2^{L+1}AN^2 - 2\nu N^2}{2^{3L+1}(A + \nu N)^2} + \frac{\nu N^3((2L + 1)A + 2(2L - 1)\nu + 3\nu N)}{2^{3L+2}(A + \nu N)^3} \right)
\]
\[
- \frac{\nu N^2}{2^{3L+1}(A + \nu N)}
\]

The population of bacteria with spacer \(\alpha\) in the second position of the spacer array \(n_{\cdot,\alpha} = \sum_{\alpha'} n_{\alpha',\alpha}\). The variance is defined as
\[
\langle n_{\cdot,\alpha}^2 \rangle = \sum_{\alpha'} \langle n_{\alpha',\alpha}^2 \rangle + \sum_{\langle \alpha'\beta' \rangle} \langle n_{\alpha',\beta'}n_{\beta'\alpha} \rangle
\]
\[
= (2L - 1)\langle n_{3\alpha}^2 \rangle + (2L - 1)(2L - 2)\langle n_{\alpha\alpha}n_{\gamma\alpha} \rangle + 2(2L - 1)\langle n_{\alpha\alpha}n_{3\alpha} \rangle
\]
\[
= \frac{N^2}{2^{3L+1}(A + N\nu)^3} \left( 2^{1+L}A^3 + 2A^2\nu(-1 + 2L + N + 2^{1+L}N) \right.
\]
\[
+ 3AN\nu^2(-1 + 2L + N + 2^LN) + 2N^2\nu^3(-1 + 2L + N) \right]
\]

### 3.6.2 Phage diffusion

By the Stokes-Einstein result, \(D = \frac{kT}{6\pi\mu r}\), where \(\mu\) is the viscosity, and \(r\) is the radius of particles. For water, \(\mu = 0.653 \times 10^{-3}\text{kg/(m}\cdot\text{s})\) at 40°C, \(\mu = 0.547 \times 10^{-3}\text{kg/(m}\cdot\text{s})\) at 50°C, \(\mu = 0.467 \times 10^{-3}\text{kg/(m}\cdot\text{s})\) at 60°C [57].
The diffusion equation is
\[ \frac{\partial n}{\partial t} = D \nabla^2 n, \] (3.13)
where \( n \) is the density of particles. By conservation of particles,
\[
n(r, t + \Delta t) = a \cdot n(r + \Delta x, t) + a \cdot n(r - \Delta x, t) \\
+ a \cdot n(r + \Delta y, t) + a \cdot n(r - \Delta y, t) \\
- 4a \cdot n(r, t),
\]
where \( a \) is the probability for each particle jumping to the near site. Expanding the right hand side and left hand side, one obtains
\[ \frac{\partial n}{\partial t} \Delta t = a l^2 \nabla^2 n \] (3.14)
Comparing Eq. 3.13 and Eq. 3.14,
\[ \frac{a}{\Delta t} \frac{D}{l^2} = \frac{kT}{6\pi \mu r l^2}. \] (3.15)
The length of each site, \( l \), should be the same order of the size of bacteria. For E. coli, the typical length is 1\( \mu \text{m} \sim 4 \mu \text{m} \). We use \( l = 4 \mu \text{m} \). The radius of a phage is the order of 100nm. We set \( T = 333 \text{K}(60^\circ \text{C}) \). Thus, we find the value of \( \frac{\Delta}{\Delta t} = 0.326 \text{min}^{-1} \).

If we consider the cells around the phage, the effective diffusion coefficient \( D_{\text{eff}} = \frac{2-2\phi}{2+\phi} D \) [58], where \( \phi \) is the volume fraction of bacteria in the lattice. If \( \phi = 0.2 \), then \( \frac{D_{\text{eff}}}{D} = 0.73 \), and \( D_{\text{eff}} \approx 0.2l^2 \cdot \text{min}^{-1} = 3.2 \mu \text{m}^2 \cdot \text{min}^{-1} \).

Even in a neutral model, the spacer diversity decreased from the leader end to the distal end, Fig. 3.14a. Fig. 3.18 shows that when the bacterial population became very large, the decrement of the diversity was almost zero. This stochastic reduction of diversity is much smaller than that induced by selection in all but the very smallest populations.

**Figure 3.18:** In the neutral model, when the population of bacteria was very large, the difference between the spacer’s diversity in different positions was almost 0. Here \( N = 500000 \), and other parameters are the same as those in Fig. 3.14.

Sensitivity of model results to number of proto-spacers per virus is shown in Fig. 3.20 and Fig. 3.21. From Fig. 3.21, the spacer turnover is larger during the first
Figure 3.19: **Daily abundance of the 10-most-abundant spacers.** (a) The phage diffusion constant is $32 \mu m^2 \cdot min^{-1}$, (b) The phage diffusion constant is $0.32 \mu m^2 \cdot min^{-1}$. Other parameters are the same as those in Fig. 3.4.

several days when phages have more proto-spacers, in agreement with experiment, supporting our conclusion that multiple proto-spacers are necessary to describe these experimental data. These results show that multiple proto-spacers are necessary to capture the experimental spacer turnover data.

Figure 3.20: **Sensitivity of spacer turnover to different number of proto-spacers per phage.** Other parameters are the same as those in Fig. 3.4.

The sensitivity of spacer turnover to spacer incorporation rate and phage mutation rate was examined. In Fig. 3.22, probabilities of acquiring new spacers are $\gamma = \ldots$
Figure 3.21: **Sensitivity of spacer abundance to different number of proto-spacers per virus.** (a) There is one proto-spacer in phage genome. (b) There are 10 proto-spacers in phage genome. (c) There are 20 proto-spacers in phage genome. (d) There are 30 proto-spacers in phage genome. (e) There are 40 proto-spacers in phage genome. (f) There are 50 proto-spacers in phage genome. Other parameters are the same as those in Fig. 3.4.

0.01 − 0.1. In Fig. 3.23, mutation rates of proto-spacer are $\mu = 0.001 − 0.1$ per generation per proto-spacer. The behaviors of spacer turnover in these two figures are qualitatively the same, which mean that the spacer turnover is robust for ranges of these two parameters examined here.

Because initially bacteria have no spacers, phages experience exponential growth and almost reach the maximum population size. As bacteria incorporated new spacers, the phage population can decrease, depending on the mutation rate and the probability of acquiring new spacers. Thus, there can be a peak in the phage pop-
Figure 3.22: **Spacer turnover for different probabilities of acquiring new spacers.** Probabilities of acquiring new spacers are $\gamma = 0.01 - 0.1$. Other parameters are the same as those in Fig. 3.4.

Figure 3.23: **Spacer turnover for different mutation rates of proto-spacers.** Mutation rates of proto-spacer are $\mu = 0.001 - 0.1$ per generation per proto-spacer. Other parameters are the same as those in Fig. 3.4.

ulation. Fig. 3.24 shows that for a higher probability of acquiring new spacers, the decrease of phage population was more dramatic.
Figure 3.24: **The population of phages for different probabilities of acquiring a new spacer.** Phages were placed on a strip initially. The population of phages in the area near the strip was tracked. The probabilities of acquiring a new spacer are $\gamma = 0.005 - 0.1$, mutation rate is $\mu = 0.0002$ per generation per proto-spacer, number of spacers is one, number of proto-spacer is one, and other parameters are the same as those in Fig. 3.6. Error bars are one standard error.

Fig. 3.25 shows a fast initial growth of bacteria with spacers. This is because the bacteria that gained spacers had a larger fitness, and they eventually dominated the bacterial population.

From Fig. 3.26, one can observe that increasing the number of proto-spacers made the phage more susceptible to the CRISPR immune system.

Fig. 3.27 shows the population of bacteria and phages for different number of spacers. Results for larger number of spacers are shown in Fig. 3.28. Both Fig. 3.27 and Fig. 3.28 show that increasing the number of spacers led to a stronger CRISPR immunity.

We do a sensitivity analysis of the effect of varying the infection probability, $S$, by examining $\langle S_{\text{recombination}} \rangle - \langle S_{\text{point mutation}} \rangle$, between the recombination and mutation, Fig. 3.29 and Fig. 3.30. A phase graph, Fig. 3.31, shows that recombination is more
Figure 3.25: The fraction of bacteria with spacers in the total bacteria population. Parameters are the same as those in Fig. 3.4. A fast initial growth of bacteria with spacers can be observed.

Figure 3.26: The population of bacteria when there was one proto-spacer locus and 30 proto-spacer loci. Other parameters are the same as those in Fig. 3.6. Error bars are one standard error.

effective than point mutation for a wide range of values of probability of acquiring new spacer and proto-spacer mutation rates. In Fig. 3.31, the color represents the
Figure 3.27: The population of (a) bacteria and (b) phages when there was one spacer or four spacers. The number of proto-spacers is one, probability of acquiring a new spacer is $\gamma = 0.1$, mutation rate of phages is $\mu = 0.0001$ per generation per proto-spacer, and other parameters are the same as those in Fig. 3.4. Error bars are one standard error.

Figure 3.28: The population of (a) bacteria and (b) phages when there was one spacer or 30 spacers. The number of proto-spacers is two, mutation rate of phages is $\mu = 0.001$ per generation per proto-spacer, and other parameters are the same as those in Fig. 3.27. Error bars are one standard error.

We also do a sensitivity analysis of the spacer turnover. Fig. 3.32 shows the effect of varying $\delta$. The results are similar for the range of $\delta$ shown. To show the difference of Fig. 3.5 and Fig. 3.4A, we show these two figures together in Fig. 3.33.

We enlarged the system size to $80 \times 80$ to study of system size dependence. We recalculated Fig. 3.4. Fig. 3.34 shows that, within the noise, the results are the same.
Figure 3.29: The difference in phages not recognized by CRISPR for recombination or point mutation dynamics, $\langle S_{\text{recombination}} \rangle - \langle S_{\text{point mutation}} \rangle$ for different probabilities of acquiring new spacers. Probabilities of acquiring new spacers are $\gamma = 0.01 - 0.1$. Other parameters are the same as those in Fig. 3.4.

Figure 3.30: The difference in phages not recognized by CRISPR for recombination or point mutation dynamics, $\langle S_{\text{recombination}} \rangle - \langle S_{\text{point mutation}} \rangle$ for different proto-spacer mutation rate. Mutation rates of proto-spacer are $\mu = 0.001 - 0.1$ per generation per proto-spacer. Other parameters are the same as those in Fig. 3.4.
Figure 3.31: Different of the infection probability, $S$, between the recombination and mutation for different probabilities of acquiring new spacers and mutation or recombination rates. The color represents the sum of $\langle S_{\text{recombination}} \rangle - \langle S_{\text{point mutation}} \rangle$ over time in each curve shown in Fig. 3.29 and Fig. 3.30. Probabilities of acquiring new spacers are $\gamma = 0.01 - 0.1$. Mutation or recombination rates of proto-spacer are $\mu, \nu = 0.001 - 0.01$ per generation per proto-spacer. There are two types of phages initially. Other parameters are the same as those in Fig. 3.13.

Fig. 3.35 shows the frequency of newly acquired spacer types in the second position of the spacer array during the first 10 days. This figure is a comparison Fig. S4(B) from [41]. Fig. 3.36 is the daily abundance of most 10 abundant spacers and spacer turnover in the first 10 days when the initial MOI is 10. This figure is a comparison to Fig. S2 from [41].
Figure 3.32: Spacer turnover for different phage decay rates. Phage decay rates are $\delta = 0.005 - 0.025 \text{ min}^{-1}$. Other parameters are the same as those in Fig. 3.4.

Figure 3.33: The comparison of Fig. 3.5 and Fig. 3.4A.
Figure 3.34: (a) Daily abundance of the 10-most-abundant spacers. (b) Spacer turnover during the first 10 days. The grid size is $L = 80$. Other parameters are the same as those in Fig. 3.4.

Figure 3.35: Frequency of newly acquired spacers types in the position +2 in the first 10 days. Here the number of spacers in CRISPR is two. Other parameters are the same as those in Fig. 3.4.
Figure 3.36: (a) Daily abundance of the most 10 abundant spacers. (b) Spacer turnover during the first 10 days. Here MOI is 10. Other parameters are the same as those in Fig. 3.4.
Chapter 4

Nonclassical Phase Diagram for Virus Bacterial Co-evolution Mediated by CRISPR

In this chapter, we find an interesting phase diagram of the extinction probability of phages, which cannot be explained by the classical predator-prey model. In the classical predator-prey model, bacteria and phages only coexist within one parameter region. Outside this region, bacteria and phages cannot both coexist. In this paper, we find bacteria and phages can coexist in several parameter regions. The bacteria and phages coexistence is re-entrant as a function of the exposure rate of phages to bacteria, for low phage mutation rates.

4.1 Methods

We used a stochastic model to study the population dynamics of bacteria and phages. The bacteria have a rate of acquiring and losing spacers. The phages have multiple proto-spacers that can evolve by point mutation and recombination. Spacers and proto-spacers are expressed as a bit string. Each bit can be either “0” or “1”. The length of each spacer and proto-spacer is \( L \) bits. The number of proto-spacers in phages is \( n_p \). CRISPR suppresses the phages, and unrecognized phages can infect and reproduce in bacteria. The coevolving dynamics is described by seven events:

1. **Bacteria reproduction**: The growth rate of wild type bacteria that do not acquire any spacers is \( c_0 \). Each spacer has a cost \( c \). The growth rate of bacteria
that have spacer array $\vec{s}$ is $[1 - (x^B + x^I)/x_{mB}] \cdot c_0/(1 + c \cdot k_{s})$, where $k_s$ is the number of spacers in the spacer array $\vec{s}$, $x^B$ is the density of healthy bacteria, $x^I$ is the density of infected bacteria, and $x_{mB}$ is the maximum density of bacteria.

2. **Bacteria infection**: Healthy bacteria can be infected by phages. The adsorption rate of phages to healthy bacteria is $\beta x^P x^B$, where $\beta$ is the exposure rate, $x^P$ is the density of free phages, and $x^B$ is the density of healthy bacteria. Bacteria have a probability $\gamma$ to acquire a new spacer from the invading phage genome. Each proto-spacer has probability $\gamma/n_p$ to be acquired. The newly acquired spacer is always inserted at the leader-proximal end of CRISPR, and the phage is degraded. Old spacers are shifted to the distal end. The maximum number of spacers per bacteria is $n_s$. If the number of spacers reaches $n_s$, the oldest spacer is deleted when a new spacer is acquired. The alternative event, with probability $1 - \gamma$, is no incorporation of a proto-spacer. In this case, if any spacer in the CRISPR matches any proto-spacer of the phage, the phage is killed. Otherwise, this bacterium becomes infected.

3. **CRISPR deletes one spacer**: A bacterium that possesses the spacer array $\vec{s}$ has a rate $d \cdot k_{s}$ to delete one spacer, where $k_{s}$ is the number of spacers in spacer array $\vec{s}$, and $d$ is the rate of deleting one spacer. When one spacer is deleted, the other spacers will be shifted towards the leader end.

4. **Bacterial lysis**: Infected bacteria have a rate $1/\tau$ to lyse, where $\tau$ is the latent time. When the infected bacteria lyse, $b$ new phages come out, where $b$ is the burst size. Each of the newborn phages can have point mutations or recombination.

5. **Phage mutation**: Phages upon bacterial lysis can have point mutation. The
rate of point mutation is $\mu$ per base per replication. A mutation flips the value of a nucleotide.

6. **Phage recombination**: Phages upon bacterial lysis can have recombination. The rate of recombination is $\nu$. A recombination occurs with another phage randomly in the whole phage population, as an approximation to multiple infection. The recombination crossover probability is $p_c$.[27]

7. **Phage degradation**: Each phage has a decay rate $\delta$.

Initially, no bacteria have spacers. There are one or more strains of phages in the environment initially. Each strain of phages has $n_p$ distinct proto-spacers.

The values of parameters are determined by the experimental data (see Supplementary Information). We used the Lebowitz-Gillespie algorithm [53, 54] to sample the stochastic process of the coevolution of bacteria and phages. The master equation of this stochastic process is in the Appendix.

### 4.2 Results

Fig. 4.1 shows a phase diagram for the phage and bacterial populations. In Fig. 4.1(c), there are four transitions in the extinction probability of phages when the mutation rate of phages is small. In region (1), phages begin to emerge but the density of phages stabilizes at a low level. Bacteria and phages can coexist in this region. In region (2), the density of phages increases initially but then decreases to zero. In this region, phages have a high probability to go extinct. In region (3), the density of phages initially increases and then decreases, but in contrast to the behavior in region (2), phages can grow back and avoid extinction in this case. In region (4), phages rapidly go extinct after a sharp initial burst. The extinction probability of phages is
Figure 4.1: The extinction probability of bacteria and phages when $\gamma = 0.0005$ and $d = 10^{-5}$ min$^{-1}$. The values of other parameters are $c_0 = 0.005$ min$^{-1}$, $c = 0.1$, $b = 100$, $\tau = 40$ min, $\delta = 0.001$ min$^{-1}$, $\nu = 0$, $n_p = 30$, $L = 10$, $n_s = 6$, and $x_B = 10^7$ mL$^{-1}$. The volume of the system in our simulation is $V = 10^{-3}$ mL. There is one strain of phages initially. The initial bacterial density is $5 \times 10^6$ mL$^{-1}$. The initial phage density is $5 \times 10^7$ mL$^{-1}$. (a) The extinction probability of phages. (b) The extinction probability of bacteria. (c) The extinction probability of phages when $\mu = 10^{-8}$ per base per replication in Fig. 4.1(a). (d) The typical behavior of a density of phages in the first four regions of Fig. 4.1(c). (e) The average number of spacers in bacteria in the first four regions of Fig. 4.1(c). (f) The average number of spacers in bacteria when bacteria go extinct and when phage go extinct.
high, and the extinction probability approaches a limit. In this region, bacteria and phages cannot coexist.

The four transitions for the extinction probability of phages as a function of $\beta$ can be explained by Eq. 4.1 and Eq. 4.2. In region (1), because the density of phages is low and the value of $\beta$ is small, the number of spacers in bacteria is almost 0, Fig. 4.1(e). Therefore, almost all bacteria are susceptible to phages. The equations of infected bacteria and phages can be approximated as

$$\frac{dx^I}{dt} = \beta x^P x^B - \frac{x^I}{\tau} \quad \text{(4.1a)}$$
$$\frac{dx^P}{dt} = \frac{b}{\tau} x^I - \beta x^P x^B - \delta x^P \quad \text{(4.1b)}$$

where $x^I$ is the density of infected bacteria and $x^P$ is the density of phages. Solving Eq. 4.1, we find when $\beta^* = \delta/[x^B(b-1)] \approx 10^{-12}$ mL $\cdot$ min$^{-1}$ the replication rate of phages begins to exceed the phage decay rate, so phages emerge in the system.

In region (2), as $\beta$ increases, the density of phages rapidly increases and bacteria begin to acquire spacers, Fig. 4.1(d) and (e). We can estimate the selection pressure on bacteria in this region. When $x^P \approx 10^9$ mL$^{-1}$, which is the typical density of phages before bacteria acquire spacers in region (2), the infection rate of each bacterium that has no spacers is $\beta x^P \approx 10^{-3}$ min$^{-1}$, which is the same order as the growth rate of bacteria. So the bacteria that acquire spacers dominate the bacterial population in a short time, and the density of phages will go down, eventually to zero.

In region (3), the phages increase first, then bacteria acquire spacers, leading the phages to decrease, which is similar to the behavior in the region (2). But when the density of phages is low, bacteria will delete spacers due to the deletion rate and the cost of spacers. Because the mutation rate of phages is small, bacteria that acquire one or more spacers have immunity against most phages. Phages can only infect
those bacteria that lost all spacers. Here we define the proportion of bacteria that have lost all spacers as $q$. Then the density of susceptible bacteria is $x^B \cdot q$. Thus the equation of infected bacteria can be approximated as

$$\frac{dx^I}{dt} = \beta x^P x^B q - \frac{x^I}{\tau}$$

(4.2)

In region (3), the value of $q$ is roughly 0.1 from Fig. 4.1(e), so we can find $\beta^* = \delta/[x^B(qb - 1)] \approx 10^{-11}$ mL · min$^{-1}$. Therefore, in region (3), phages can grow back when some portion of bacteria lose spacers. As the density of phages increases, the average number of spacers in bacteria also increases, which in turn represses the growth of phages, as in Figs. 4.1(d) and (e). So in this case, the density of phages fluctuates around a low value and eventually stabilizes.

The density of free phages decreases due to two factors. One is decay. The other one is due to CRISPR recognition and subsequently degradation. Therefore, the overall decay rate of phages is $\beta x^B + \delta$. In the left boundary of region (4), $\beta$ is the order of $10^{-9}$ mL · min$^{-1}$, and the overall decay rate of phages is $\beta x^B + \delta \approx 10^{-2}$ min$^{-1}$.

Following the same argument as in region (3), the minimum value of $q$ for which phages can grow back is $q^* = (\beta x^B + \delta)/(b\beta x^B) \approx 10^{-2}$. The time required for $q^*$ bacteria to lose spacers is $t > q^*/d \approx 1000$ min, which is longer than the half life of phages. Thus, before bacteria lose spacers, all of phages are adsorbed into bacteria. Because bacteria have acquired spacers and the mutation rate of phages is small, phages that are adsorbed into bacteria are killed by CRISPR. Therefore, in region (4), phages go extinct rapidly after the initial burst. When $\beta$ further increases, if bacteria acquire spacers, phages will go extinct rapidly. If bacteria do not acquire spacers, bacteria will go extinct, as in Fig. 4.1(f). The extinction probability of phage approaches a limit, $1 - (1 - \gamma)^{N_0^B} \approx 0.918$ in Fig. 4.1, the probability that one of the initial bacteria acquired a spacer, where $N_0^B$ is the initial bacterial population.
Figure 4.2: The extinction probability of (a) phages and (b) bacteria. Here $\gamma = 0.0005$ and $d = 0.0001 \text{ min}^{-1}$. The extinction probability of (c) phages and (d) bacteria. Here $\gamma = 0.005$ and $d = 10^{-5} \text{ min}^{-1}$. Other parameters are the same as those in Fig. 4.1.

From the above explanation of the four regions in Fig. 4.1, we have the conditions for which this interesting non-classical phase diagram of phage extinction exists. First, bacteria must possess the CRISPR adaptive immune system: if bacteria do not have CRISPR, bacteria and phages can only coexist when $\beta$ is small, $\beta \approx 10^{-12} \text{ mL} \cdot \text{ min}^{-1}$, and region (2) and region (4) will not exist. Second, bacteria must have some rate to lose the acquired immunity. If bacteria can accumulate an unlimited number of spacers, phages will eventually go extinct if the length of the proto-spacers is finite and region (3) will not exist. Third, the rate of losing the adaptive immunity must be small. In region (2) and the left boundary of region (4), phages cannot grow back because the rate of losing spacers is small. If the rate of losing spacers is large,
region (2) will disappear and the left boundary of region (4) will move towards higher $\beta$ values, as shown in Fig. 4.2(a) and (b). Conversely this phase diagram is not sensitive to the probability of acquiring new spacers. Increasing $\gamma$ only changes the pattern of the extinction probability in high $\mu$ regions, making it more difficult for phages to escape from CRISPR recognition, as shown in Fig. 4.2(c) and (d). From the above results, we predict when the deletion rate of spacer and the mutation rate of phages is small, decreasing the adsorption rate of phages can make phages extinct. However, further decreasing the adsorption rate can allow phages to reemerge.

Figure 4.3 : The diversity of the first spacer, defined as $D = - \sum_k p_{s_0}(k) \log p_{s_0}(k)$, where $p_{s_0}(k)$ is the observed probability to have sequence $k$ at the first position, $s_0$, in the spacer array in CRISPR, for different values of $\beta$ when $\mu = 10^{-8}$. The other parameters are the same as those in Fig. 4.1.

CRISPR changes the bacterial population structure. In Fig. 4.3, the Shannon entropy of the first spacer is used as a measure of the diversity. In Fig. 4.3, the diversity of spacers rises slowly when $\beta$ is small, region (1) in Fig. 4.1(c). This is because the selection pressure on bacteria is small, and CRISPR does not provide
bacteria much advantage. As $\beta$ increases, the diversity of spacers increases faster because the density of phages is larger and the value of $\beta$ is higher, making the adsorption of phages into bacteria more rapid. But the steady value of the diversity decreases, implying the distribution of spacers becomes more biased. If the selection pressure on bacteria is larger, the bacteria that acquire spacers will dominate the population in a shorter time. When the bacteria that have spacers dominate the population, phages are repressed, and the density of phages stays at a low level. The process of acquiring spacers becomes slower, leading to a smaller steady value of spacer diversity.

Figure 4.4: The extinction probability of (a) phages and (b) bacteria when phages undergo only point mutation. The extinction probability of (c) phages and (d) bacteria when phages undergo only recombination. Initially, there are two strains of phages. Here $\gamma = 0.01$. The other parameters are the same as those in Fig. 4.1.

Recombination is compared to point mutation of phages in Fig. 4.4. Here there are two strains of phages initially, and so acquisition of two spacers is required for bacterial immunity. The limiting extinction probability in this case is $1 - (1 - \gamma^2)^{N^B} \approx 0.40$ in Fig. 4.4. Additionally, at very large $\beta$, bacteria with a finite number of spacers,
eventually go extinct when the spacer array by chance is entirely occupied by proto-spacers from only one strain of phage. Finally, the extinction probability of phages when phages have only recombination is lower than that when phages have only point mutation, because recombination can change several proto-spacers at once.

4.3 Discussion

In summary, we predict an interesting phase diagram of phage extinction probability. When the deletion rate of spacers in CRISPR is small, phages go extinct when the value of $\beta$ is low, but phages can coexist with bacteria when $\beta$ is even lower. CRISPR changes the evolution of bacteria and phages, accelerating the coevolution of bacteria and phages. Finally, recombination is a more efficient mechanism for phages to escape the recognition of CRISPR than is point mutation when there are multiple proto-spacers in the phage. Future work may consider biotechnology applications, genome editing approaches, population-level bacterial control, or effects of recombination in the microbiome.

4.4 Appendix

4.4.1 Table of parameters

The parameters used in the main text are listed in table 4.1.

The values of $c_0$ and $x^B_M$ are estimated from [41]. The cost of spacers is low [40]; here we choose $c = 0.1$. The values of $b$ and $\tau$ are estimated from [56]. The value of $\beta$ is estimated from [59]. The value of $\gamma$ is estimated from [28] and [35]. The value of $d$ is estimated from [40]. The value of $\delta$ is estimated from [60]. The value of $\mu$ is estimated from [61]. The value of $\nu$ is estimated from [62] and is the same order as the
Table 4.1: Table of parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Value</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_0$</td>
<td>Bacterial growth rate</td>
<td>$0.005 \text{ min}^{-1}$</td>
<td>[41]</td>
</tr>
<tr>
<td>$c$</td>
<td>Cost of each spacer</td>
<td>0.1</td>
<td>[40]</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Bacterial exposure rate</td>
<td>$10^{-13} - 10^{-5} \text{ mL \cdot min}^{-1}$</td>
<td>[59]</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Probability of acquiring new spacers</td>
<td>0.0005</td>
<td>[28], [35]</td>
</tr>
<tr>
<td>$d$</td>
<td>Rate of deleting one spacer</td>
<td>$10^{-2} \text{ min}^{-1}$</td>
<td>[40]</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Latent time</td>
<td>40 min</td>
<td>[56]</td>
</tr>
<tr>
<td>$b$</td>
<td>Phage burst size</td>
<td>100</td>
<td>[56]</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Phage decay rate</td>
<td>0.001 $\text{ min}^{-1}$</td>
<td>60</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Mutation rate</td>
<td>$10^{-10} - 10^{-2}$ per base per replication</td>
<td>61</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Recombination rate</td>
<td>$10^{-10} - 10^{-2}$ per base per replication</td>
<td>62</td>
</tr>
<tr>
<td>$n_p$</td>
<td>Number of proto-spacers in phages</td>
<td>30</td>
<td>[41]</td>
</tr>
<tr>
<td>$n_s$</td>
<td>Maximum number of spacers in CRISPR</td>
<td>6</td>
<td>[41]</td>
</tr>
<tr>
<td>$L$</td>
<td>Length of each spacer and proto-spacer</td>
<td>10 bp</td>
<td>[63]</td>
</tr>
<tr>
<td>$x_M^B$</td>
<td>Maximum bacterial density</td>
<td>$10^9 \text{ mL}^{-1}$</td>
<td>[41]</td>
</tr>
<tr>
<td>$V$</td>
<td>Volume of the system</td>
<td>$10^{-3} \text{ mL}^{-1}$</td>
<td></td>
</tr>
</tbody>
</table>

The interference between proto-spacers and CRISPR spacers is governed by the PAM and the seed regions [63]. The length of the PAM is about 3 bp and the length of the seed region is 7 bp [63], so we set the length of spacers and proto-spacers to 10 pb. In the experiment to which we compare [41], the average number of spacers in CRISPR is small, on average 0.8 spacers per bacteria, so we set the maximum number of spacers to 6. The average number of spacers in our simulation is shown in Fig. 4.5. There are 27 spacers that account for between 82% and 99% of all spacers sampled on any individual day in the experiment to which we compare [41]. Here we set $n_p$ to 30. We also tried $n_p = 1$. The results are qualitatively the same, as shown
in Fig. 4.7. The volume $V$ is set to mimic the typical volume of a droplet.

4.4.2 Master equation

The master equation of the stochastic process described in the main text is
\[
\frac{dP\{N^B_s, N^I_{\bar{p}_1}, N^P_{\bar{p}_2}\}}{dt} = \sum_{\bar{s}} \frac{c_0}{1 + c \cdot k_{\bar{s}}} (N^B_{\bar{s}} - 1) \left\{ 1 - \left[ \left( \sum_{\bar{s}'} N^B_{\bar{s}'} \right) - 1 + \sum_{\bar{p}} N^I_{\bar{p}} \right] / N^B_M \right\} \\
P(N^B_s - 1) - \sum_{\bar{s}} \frac{c_0}{1 + c \cdot k_{\bar{s}}} N^B_{\bar{s}} \left[ 1 - \left( \sum_{\bar{s}'} N^B_{\bar{s}'} + \sum_{\bar{p}} N^I_{\bar{p}} \right) / N^B_M \right] P(N^B_{\bar{s}}) \\
+ \frac{(1 - \gamma)}{V} \sum_{\bar{s}} \sum_{\bar{p}_2} (N^P_{\bar{p}_2} + 1)(N^B_{\bar{s}} + 1) \theta_1(\bar{p}_2, \bar{s}) P(N^B_{\bar{s}} + 1, N^I_{\bar{p}_2} - 1, N^P_{\bar{p}_2} + 1) \\
+ \frac{(1 - \gamma)}{V} \sum_{\bar{s}} \sum_{\bar{p}_2} (N^P_{\bar{p}_2} + 1) N^B_{\bar{s}} \left[ 1 - \theta_1(\bar{p}_2, \bar{s}) \right] P(N^B_{\bar{s}}, N^P_{\bar{p}_2} + 1) \\
- \frac{\beta}{V} \sum_{\bar{s}} \sum_{\bar{p}_2} \sum_{\bar{p}} N^P_{\bar{p}_2} N^B_{\bar{s}} P(N^B_{\bar{s}}, N^P_{\bar{p}_2}) + \frac{\gamma \beta}{V \cdot N_p} \\
\sum_{\bar{s}} \sum_{\bar{s}'} \sum_{\bar{p}_2} \sum_{\bar{p}} (N^P_{\bar{p}_2} + 1)(N^B_{\bar{s}'} + 1) \theta_2(\bar{p}_2, \bar{s}', \bar{s}) P(N^B_{\bar{s}'} + 1, N^B_{\bar{s}} - 1, N^P_{\bar{p}_2} + 1) \\
+ d \sum_{\bar{s}} \sum_{\bar{s}'} (N^B_{\bar{s}'} + 1) \sum_{i=1}^{\bar{k}_{\bar{s}'}} \theta_3(s_i, \bar{s}', \bar{s}) P(N^B_{\bar{s}'} + 1, N^B_{\bar{s}} - 1) - d \sum_{\bar{s}} k_{\bar{s}} N^B_{\bar{s}} P(N^B_{\bar{s}}) \\
+ \delta \sum_{\bar{p}_2} (N^P_{\bar{p}_2} + 1) P(N^P_{\bar{p}_2} + 1) - \delta \sum_{\bar{p}_2} N^P_{\bar{p}_2} P(N^P_{\bar{p}_2}) \\
+ \frac{1 - \nu}{\tau} \sum_{\bar{p}_1} \sum_{\bar{p}'_1} \sum_{\bar{p}'_2} \sum_{\bar{p}_3} (N^I_{\bar{p}_1} + 1) \prod_{i=1}^{b} \left( \mu^b(\bar{p}_1, \bar{p}'_i)(1 - \mu)^{L-n_p-h(\bar{p}_1, \bar{p}'_i)} \right) \\
\cdot P(N^I_{\bar{p}_1} + 1, N^P_{\bar{p}_1} - 1, \ldots, N^P_{\bar{p}_b} - 1) + \frac{\nu}{\tau[(\sum_{\bar{p}_2} N^P_{\bar{p}_2}) - b]^b} \\
\sum_{\bar{p}_1} \sum_{\bar{p}'_1} \sum_{\bar{p}'_2} \sum_{\bar{p}_3} (N^I_{\bar{p}_1} + 1) \prod_{i=1}^{b} \left[ \mu^b(\bar{p}_1, \bar{p}'_i)(1 - \mu)^{L-n_p-h(\bar{p}_1, \bar{p}'_i)} \right] \\
\cdot (N^P_{\bar{p}_1} - \sum_{j=1}^{b} \Delta_{\bar{p}_1, \bar{p}'_j}) \sum_{\bar{r}} \prod_{k=1}^{L-n_p} \bar{p}_n^{\bar{r}_k} \left( 1 - p_c \right)^{1-r_k} \theta_4(\bar{p}'_1, \bar{p}_1, \bar{p}'_2, \bar{p}_2, \bar{r}) \right] \\
\cdot P(N^I_{\bar{p}_1} + 1, N^P_{\bar{p}_1} - 1, \ldots, N^P_{\bar{p}_b} - 1) - \frac{1}{\tau} \sum_{\bar{p}_1} N^I_{\bar{p}_1} P(N^I_{\bar{p}_1}) \\
\tag{4.3}
\]

where $N^B_s$ is the population of the bacteria with spacer array $\bar{s}$, $N^I_{\bar{p}_1}$ is the population of infected bacteria invaded by phages with proto-spacer array $\bar{p}_1$, $N^P_{\bar{p}_2}$ is the population of phages with proto-spacer array $\bar{p}_2$, and $N^B_M$ is the maximum popula-
tion of bacteria. In Eq. 4.3, \( \theta_1(\vec{p}, \vec{s}) = 0 \) when \( \vec{s} \) recognizes \( \vec{p} \) and 1 otherwise. The \( \theta_2(p_i, s', \vec{s}) = 1 \) when \( \{p_i, s'_1, \ldots, s'_{n_s-1}\} = \vec{s} \) and 0 otherwise. The \( \theta_3(s'_i, s'', \vec{s}) = 1 \) when \( \{s'_1, \ldots, s'_{i-1}, s'_{i+1}, \ldots, s'_{n_s}, 0\} = \vec{s} \) and 0 otherwise. The hamming distance between \( \vec{p}_i \) and \( \vec{p}'_i \) is \( h(\vec{p}_i, \vec{p}'_i) \). The \( \vec{r} \) is a bit string, which denotes the recombination pattern.

Each bit in \( \vec{r} \) is either 0 or 1. If \( \vec{r}_k = 1 \), it means there is a crossover at position \( k \). If \( \vec{r}_k = 0 \), it means there is no crossover at position \( k \). The \( \theta_4(\vec{p}'_i, \vec{p}^*_i, \vec{p}''_i, \vec{r}) = 1 \) if \( \vec{p}''_i \) can be generated by the recombination pattern \( \vec{r} \) from \( \vec{p}'_i \) and \( \vec{p}^*_i \) and 0 otherwise.

\( \Delta_{\vec{p}_i, \vec{p}'_j} = 1 \) if \( \vec{p}_i = \vec{p}'_j \) and 0 otherwise. In \( P(N^l_{\vec{p}_i} + 1, N^p_{\vec{p}^*_i} - 1, \ldots, N^p_{\vec{p}'_b} - 1) \), if \( \vec{p}_i = \vec{p}'_j \), it means \( N^p_{\vec{p}'_k} - 2 \). In general, \( P(N^l_{\vec{p}_i} + 1, N^p_{\vec{p}^*_i} - 1, \ldots, N^p_{\vec{p}'_b} - 1) \) is short hand for \( P(N^l_{\vec{p}_i} + 1, \sum_{j=1}^{b} \Delta_{\vec{p}_k, \vec{p}'_j} \}) \).

We can show that

\[
\sum_{\vec{p}'_1, \ldots, \vec{p}'_b} \prod_{i=1}^{b} \left( \mu^h(\vec{p}_i, \vec{p}'_i) (1 - \mu)^{L-n_p - h(\vec{p}_i, \vec{p}'_i)} \right) = 1
\]

and

\[
\sum_{\vec{p}'_1} \sum_{\vec{r}} \sum_{k=1}^{L-n_p} \prod_{c=1}^{L-n_p} p_c^{\vec{r}_c} (1 - p_c)^{1-\vec{r}_c} \theta_4(\vec{p}'_i, \vec{p}^*_i, \vec{p}''_i, \vec{r})
\]

\[
= \sum_{\vec{r}} \prod_{c=1}^{L-n_p} p_c^{\vec{r}_c} (1 - p_c)^{1-\vec{r}_c} \sum_{\vec{p}''_i} \theta_4(\vec{p}'_i, \vec{p}^*_i, \vec{p}''_i, \vec{r})
\]

\[
= \sum_{\vec{r}} \prod_{c=1}^{L-n_p} p_c^{\vec{r}_c} (1 - p_c)^{1-\vec{r}_c}
\]

\[
= 1
\]

Therefore,

\[
\sum_{\vec{p}'_1, \ldots, \vec{p}'_b} \sum_{\vec{p}''_1, \ldots, \vec{p}''_b} \prod_{i=1}^{b} \left[ \mu^h(\vec{p}_i, \vec{p}'_i) (1 - \mu)^{L-n_p - h(\vec{p}_i, \vec{p}'_i)} \right] N^p_{\vec{p}'_b} \sum_{\vec{r}} \prod_{k=1}^{L-n_p} p_c^{\vec{r}_c} (1 - p_c)^{1-\vec{r}_c} \theta_4(\vec{p}'_i, \vec{p}^*_i, \vec{p}''_i, \vec{r}) = (\sum_{\vec{p}} N^p_{\vec{p}})^b
\]
and

\[-\frac{1 - \nu}{\tau} \sum_{\vec{p}_1} \sum_{\vec{p}_1', \ldots, \vec{p}_b'} N^I_{\vec{p}_1} \prod_{i=1}^b \left( \mu^{h(\vec{p}_1, \vec{p}_i')} (1 - \mu)^{L-n_p-h(\vec{p}_1, \vec{p}_i')} \right) P(N^I_{\vec{p}_1}) \]

\[-\frac{\nu}{\tau (N^P_{\vec{p}})^{L-n_p}} \sum_{\vec{p}_1} \sum_{\vec{p}_1'} \sum_{\vec{p}_1''} \sum_{\vec{p}_1''' \ldots \vec{p}_b''} N^I_{\vec{p}_1} \prod_{i=1}^b \left( \mu^{h(\vec{p}_1, \vec{p}_i')} (1 - \mu)^{L-n_p-h(\vec{p}_1, \vec{p}_i')} N^P_{\vec{p}_i'} \right) \]

\[\sum_{\vec{r}} \prod_{k=1}^{L-n_p} \rho(1 - p_c)^{1-r_1} \theta_1(\vec{p}, \vec{p}_1', \vec{p}_i', \vec{r}) \] \[P(N^I_{\vec{p}_1}) = -\frac{1}{\tau} \sum_{\vec{p}_1} N^I_{\vec{p}_1} P(N^I_{\vec{p}_1}) \]

This is why we get the last term in Eq. 4.3.

### 4.4.3 Mean field equations

The corresponding mean field equations for the densities of bacteria and phages, shown for illustrative purpose and not used in the main text, are

\[\frac{d x^B_s}{dt} = \frac{c_0}{1 + c \cdot k_s} x^B_s \left[ 1 - \left( \sum_{s'} x^B_{s'} + \sum_{\vec{p}} x^I_{\vec{p}}/x^B_{m1} \right) - (1 - \gamma) \beta x^B_s \sum_{\vec{p}} x^I_{\vec{p}} \theta_1(\vec{p}, \vec{s}) - \gamma \beta x^B_s \right] \]

\[\sum_{\vec{p}} x^P_{\vec{p}} + \frac{\gamma \beta}{n_p} \sum_{s'} \sum_{\vec{p}} \sum_{i=1}^{n_p} x^B_{s'} x^P_{\vec{p}} \theta_2(p_i, s', \vec{s}) - d \cdot k_s x^B_s + d \sum_{s'} x^B_{s'} \sum_{i=1}^{k_{x'}} \theta_3(s_i', s''', \vec{s}) \]

(4.4a)

\[\frac{d x^I_{\vec{p}}}{dt} = (1 - \gamma) \beta \sum_{s} x^B_s x^P_{\vec{p}} \theta_1(\vec{p}, \vec{s}) - \frac{x^I_{\vec{p}}}{\tau}, \quad (4.4b)\]

\[\frac{d x^P_{\vec{p}}}{dt} = \frac{b(1 - \nu)}{\tau} \sum_{\vec{p}'} x^I_{\vec{p}'} \mu^{h(\vec{p}, \vec{p}')} (1 - \mu)^{L-n_p-h(\vec{p}, \vec{p}')} + \frac{b\nu}{\tau} \sum_{\vec{p}'} \sum_{\vec{p}_1} \sum_{\vec{p}_2} \sum_{\vec{p}_3} x^I_{\vec{p}_1} \mu^{h(\vec{p}_1, \vec{p}_2)} \]

\[\sum_{\vec{r}} \prod_{k=1}^{L-n_p} \rho(1 - p_c)^{1-r_1} \theta_1(\vec{p}_2, \vec{p}_3, \vec{p}, \vec{r}) \]

\[\delta \cdot x^P_{\vec{p}} - \beta x^P_{\vec{p}} \sum_{s} x^B_{s}, \quad (4.4c)\]
where $x_B^{\vec{s}}$ is the density of bacteria with spacer array $\vec{s}$, $x_I^{\vec{p}}$ is the density of infected bacteria invaded by phages with proto-spacer array $\vec{p}$, and $x_P^{\vec{p}}$ is the density of phages with proto-spacer array $\vec{p}$. The functions of $\theta_1$, $\theta_2$, $\theta_3$ and $\theta_4$ are the same as those in Eq. 4.3.

### 4.4.4 Number of spacers

The average number of spacers in our simulation does not reach $n_s$ in most of the parameter regime. In the range $\beta \in [10^{-12}, 10^{-8}]$ and $\nu \in [10^{-8}, 10^{-6}]$, the average number of spacers is 0–2, which is in agreement with the experiment data in [41].

![Figure 4.5: The average number of spacers in CRISPR. The parameters are the same as those in Fig. 4.1. Blank means no data available.](image)

### 4.4.5 Number of proto-spacers

When $n_p = 1$, the pattern of the extinction probability of phages is qualitatively the same as Fig. 4.1.
Figure 4.6: The average number of spacers in CRISPR when $\mu = 10^{-8}$ in Fig. 4.5. The parameters are the same as those in Fig. 4.1.

Figure 4.7: The extinction probability of bacteria and phages when $n_p = 1$. Other parameters are the same as those in Fig. 4.1.
In this work, we developed three models to investigate the co-evolution of bacteria and phages mediated by CRISPR. These models show that CRISPR plays an important role in the bacteria and phages co-evolution. To be specific, CRISPR speeds up the evolution of bacteria and phages, making the bacteria and phages evolve faster. CRISPR changes the population structure of bacteria and phages, altering the diversity in bacteria and phage population. CRISPR also varies the conditions of the co-existence of bacteria and phages, making the phase diagram of the extinction probability of phages more complex than that of the classical predator-prey model. We also show that, from several different points of view, recombination is a more efficient strategy, than is the point mutation, for phages to escape the recognition of CRISPR.

In Chapter 2, we show that the diversity of spacers in the first position of the spacer array is higher than that of spacers closed to the distal end of the spacer array. This is because spacers record the invading history of phages into bacteria. The spacers in the first position are incorporated from the newly invaded phages. These spacers provide bacteria the highest protection against current phages. Therefore, the spacers closed to the leader end are more heterogeneous than the spacers near the distal end.

When the immunity of CRISPR requires spacers to perfectly match proto-spacers, the spacer diversity is similar when phages evolve by point mutation or recombination. It is because in this condition, a single mutation in proto-spacers is enough to lead
phages to escape the recognition of CRISPR. Therefore, recombination, which can introduce several mutations at a time, is identical in effect to point mutation.

However, point mutation and recombination have different effects when the immunity of spacers has tolerance. The protection of CRISPR on bacteria differs when one or two mismatches are required for phages to escape the recognition of CRISPR. If at least two mismatches are required for phages to avoid recognized by CRISPR, then recombination is a better strategy for phages to escape the immunity of CRISPR.

In Chapter 3, We show that an explicit, spatially-dependent model of bacteria and phages can explain the experimental data [41] on spacer abundance and turnover. We also show that when phage have multiple proto-spacer loci, recombination is a more effective escape strategy than is point mutation for phages to avoid the recognition of CRISPR. It is because recombination can change several proto-spacers at a time. But the difference between recombination and point mutation depends on the diversity of the phage population. When genetic diversity in phage population is high, recombination is almost identical to point mutation for phages to escape the recognition of CRISPR. By comparing to a neutral model, we show that the evolution of phages in this system is mainly driven by the immune pressure of bacteria imposed on phages, rather than the stochastic incorporation and shift of spacers.

Besides, we investigate the impacts of the number of spacers, the number of proto-spacers, and the phage mutation rate on the co-evolution of bacteria and phages. We predict that when the number of spacers in CRISPR increases, the diversity of phages also increases because the immune pressure upon phages enlarges. Similarly, increasing the number of proto-spacer loci also makes the immune pressure on phages increase, and the phage population decreases. Conversely, increasing the phage mutation rate allows phages to escape destruction by the immune system, and
the population of phages increases.

In Chapter 4, we allow bacteria to have a rate to lose the incorporated spacers. We find an intriguing phase diagram of phage extinction probability. When the mutation rate of phages and the deletion rate of spacers in CRISPR are small, the curve of the phage extinction probability has three peaks and two valleys. Bacteria and phages coexistence is re-entrant as a function of the exposure rate of phages to bacteria. The phase diagram of phage extinction probability is more complex than that in the classical predator-prey model. We also show that recombination is a more efficient strategy for phages to escape the immunity of CRISPR than is point mutation, using a the phase diagram to quantify the phage extinction probability.

In summary, CRISPR, which was discovered only one decade ago, challenges our understanding of the co-evolution of bacteria and phages. This work reveals several roles of CRISPR plays in the co-evolution and co-existence of bacteria and phages. The results of this work might have broad applications, such as the phage therapy and microorganism control. At last, I would like to quote one sentence of Aristotle to end this thesis, “In all things of nature there is something of the marvelous”.

Bibliography


