

# Biology of restriction-modification systems at the single-cell and population level

by

**Maroš Pleška**

October, 2017

*A thesis presented to the*

*Graduate School*

*of the*

*Institute of Science and Technology Austria, Klosterneuburg, Austria*

*in partial fulfillment of the requirements*

*for the degree of*

*Doctor of Philosophy*



*Institute of Science and Technology*



The dissertation of Maroš Pleška, titled Biology of restriction-modification systems at the single-cell and population level, is approved by:

**Supervisor:** Călin C. Guet, IST Austria, Klosterneuburg, Austria

Signature: \_\_\_\_\_

**Committee Member:** Jonathan Bollback, University of Liverpool, Liverpool, UK

Signature: \_\_\_\_\_

**Committee Member:** Bruce R. Levin, Emory University, Atlanta, USA

Signature: \_\_\_\_\_

**Exam Chair:** Gašper Tkačik, IST Austria, Klosterneuburg, Austria

Signature: \_\_\_\_\_



© by Maroš Pleška, October, 2017

All Rights Reserved

I hereby declare that this dissertation is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: \_\_\_\_\_

Maroš Pleška

December 15, 2017



# Abstract

Restriction-modification (RM) represents the simplest and possibly the most widespread mechanism of self/non-self discrimination in nature. In order to provide bacteria with immunity against bacteriophages and other parasitic genetic elements, RM systems rely on a balance between two enzymes: the restriction enzyme, which cleaves non-self DNA at specific restriction sites, and the modification enzyme, which tags the host's DNA as self and thus protects it from cleavage. In this thesis, I use population and single-cell level experiments in combination with mathematical modeling to study different aspects of the interplay between RM systems, bacteria and bacteriophages. First, I analyze how mutations in phage restriction sites affect the probability of phage escape – an inherently stochastic process, during which phages accidentally get modified instead of restricted. Next, I use single-cell experiments to show that RM systems can, with a low probability, attack the genome of their bacterial host and that this primitive form of autoimmunity leads to a tradeoff between the evolutionary cost and benefit of RM systems. Finally, I investigate the nature of interactions between bacteria, RM systems and temperate bacteriophages to find that, as a consequence of phage escape and its impact on population dynamics, RM systems can promote acquisition of symbiotic bacteriophages, rather than limit it. The results presented here uncover new fundamental biological properties of RM systems and highlight their importance in the ecology and evolution of bacteria, bacteriophages and their interactions.

# Acknowledgments

During my PhD studies, I received help from many people, all of which unfortunately cannot be listed here. I thank them deeply and hope that I never made them regret their kindness.

I would like to express my deepest gratitude to Călin Guet, who went far beyond his responsibilities as an advisor and was to me also a great mentor and a friend. Călin never questioned my potential or lacked compassion and I cannot thank him enough for cultivating in me an independent scientist. I was amazed by his ability to recognize the most fascinating scientific problems in objects of study that others would find mundane. I hope I adopted at least a fraction of this ability.

I will be forever grateful to Bruce Levin for all his support and especially for giving me the best possible example of how one can practice excellent science with humor and style. Working with Bruce was a true privilege.

I thank Jonathan Bollback and Gašper Tkačik for serving in my PhD committee and the Austrian Academy of Science for funding my PhD research via the DOC fellowship.

I thank all our lab members: Tobias Bergmiller for his guidance, especially in the first years of my research, and for being a good friend throughout; Remy Chait for staying in the lab at unreasonable hours and for the good laughs at bad jokes we shared; Anna Staron for supportively listening to my whines whenever I had to run a gel; Magdalena Steinrück for her pioneering work in the lab; Kathrin Tomasek for keeping the entropic forces in check and for her FACS virtuosity; Isabella Tomanek for always being nice to me, no matter how much bench space I took from her.

I thank all my collaborators: Reiko Okura and Yuichi Wakamoto for performing and analyzing the microfluidic experiments; Long Qian and Edo Kussell for their bioinformatics analysis; Dominik Refardt for the  $\lambda$  *kan* phage; Moritz for his help with the mathematical modeling. I thank Fabienne Jesse for her tireless editorial work on all our manuscripts.

Finally, I would like to thank my family and especially my wife Edita, who sacrificed a lot so that I can pursue my goals and dreams.

# Preface

“Le rôle de l'infiniment petit est infiniment grand.”

Louis Pasteur

Bacteria represent the oldest and most abundant forms of life on Earth and the importance of understanding them cannot be overstated. Discovery of the microbial world and the early days of microbiology were closely intertwined with the development of scientific thought itself in a period of time, during which microbes have been identified as the underlying cause of infectious diseases. Due to their relative simplicity, bacteria and their viruses (bacteriophages), later took the role of ideal model systems, where one could tackle some of the greatest problems in biology, such as uncovering the molecular basis of heredity and breaking the genetic code. However, not only are bacteria and bacteriophages the most “primitive” form of life on our planet, they are also the most abundant. Only relatively recently we have learned to appreciate that the complexity of the microbial world is not bounded by the cell envelope, and that in natural environments, bacteria and phages interact to form complex communities with principles of their own. After decades of research invested into elucidating the molecular mechanisms underlying fundamental biological processes of these simple organisms, it would seem that we are in an ideal position to comprehend their interactions as well. This thesis represents a humble attempt towards that goal.

## **About the Author**

Maroš Pleška completed a BSc in Biology at the Comenius University in Bratislava, Slovakia and an MSc in Molecular Biology, also at the Comenius University. He joined IST Austria in September 2012. His main research interests include the molecular biology of bacteria and viruses, as well as ecological and evolutionary dynamics. In 2014, he was awarded the DOC fellowship of the Austrian Academy of Sciences.

## List of publications appearing in thesis

1. Pleška M, Guet CC. (2017) Effects of mutations in phage restriction sites during escape from restriction–modification. *Biol. Lett.* 13, 20170646.
2. Pleška M, et al. (2016) Bacterial autoimmunity due to a restriction-modification system. *Curr Biol* 26(3):404–409.
3. Pleška M, et al. (2018) Phage-host population dynamics promotes prophage acquisition in bacteria with innate immunity. *Nature Ecology & Evolution* (In Press).



# Table of Contents

Abstract.....	i
Acknowledgments .....	ii
Preface .....	iii
About the Author .....	iv
List of publications appearing in thesis .....	v
Table of Contents .....	vii
List of Figures .....	ix
List of Tables.....	x
List of Symbols/Abbreviations .....	xi
<b>1 Introduction .....</b>	<b>1</b>
1.1 MOLECULAR MECHANISM AND TYPES OF RM SYSTEMS.....	1
1.2 ABUNDANCE AND MOBILITY OF RM SYSTEMS IN BACTERIA .....	3
1.3 POSSIBLE ROLES OF RM SYSTEMS.....	4
1.4 RM SYSTEMS AS A STOCHASTIC MOLECULAR MECHANISM .....	5
1.5 OTHER MECHANISMS OF PROKARYOTIC IMMUNITY .....	7
1.6 QUESTIONS ADDRESSED IN THIS THESIS.....	9
<b>2 Effects of mutations in phage restriction sites during escape from restriction- modification .....</b>	<b>11</b>
2.1 SUMMARY.....	11
2.2 INTRODUCTION .....	12
2.3 RESULTS .....	13
2.4 DISCUSSION .....	17
2.5 MATERIAL AND METHODS.....	18
<b>3 Bacterial autoimmunity due to a restriction-modification system.....</b>	<b>22</b>
3.1 SUMMARY.....	22
3.2 RESULTS .....	24
3.3 DISCUSSION .....	40
3.4 MATERIAL AND METHODS.....	42
<b>4 Phage-host population dynamics promotes prophage acquisition in bacteria with innate immunity.....</b>	<b>51</b>
4.1 SUMMARY.....	51
4.2 INTRODUCTION .....	52

4.3	RESULTS .....	54
4.4	DISCUSSION.....	75
4.5	MATERIAL AND METHODS.....	77
<b>5</b>	<b>References.....</b>	<b>94</b>

# List of Figures

Figure 1-1: The mechanism of action of RM systems.....	2
Figure 1-2: Effects of stochasticity on the function of RM systems .....	7
Figure 2-1: Genetic map of $\lambda$ cl857.....	13
Figure 2-2: Effect of mutations in restriction sites on efficiency of plating.....	15
Figure 3-1: RM plasmids and their effect on fitness.....	26
Figure 3-2: Effects of Imbalanced R and M expression .....	29
Figure 3-3: EcoRI, but not EcoRV, induces SOS response in a subpopulation of cells .....	31
Figure 3-4: Real-time dynamics of self-restriction by EcoRI in single cells .....	33
Figure 3-5: Frequency of SOS-induction and its effects growth of individual bacteria .....	34
Figure 3-6: Restriction site avoidance and enzyme specificity .....	38
Figure 3-7: Identifying determinants of self-restriction .....	39
Figure 4-1: RM systems represent a barrier to prophage acquisition in individual bacteria .	55
Figure 4-2: RM systems promote prophage acquisition at the population level .....	59
Figure 4-3: Prophage acquisition in M63 medium .....	60
Figure 4-4: Prophage acquisition in serially transferred cultures.....	61
Figure 4-5: RM systems promote prophage acquisition under a wide range of initial conditions .....	62
Figure 4-6: RM systems delay the onset of infection .....	64
Figure 4-7: Full dynamics of competition experiments in the presence of temperate phage	65
Figure 4-8: RM systems do not alter prophage acquisition in individual bacteria .....	68
Figure 4-9: Delay in the onset of infection increases prophage acquisition.....	69
Figure 4-10: Full numerical solutions of the deterministic model.....	70
Figure 4-11: Parameter sensitivity analysis for the mathematical model .....	71
Figure 4-12: Numerical solutions for different initial conditions .....	72
Figure 4-13: Effect of stochasticity on model dynamics.....	73
Figure 4-14: Model of interactions between CRISPR/Cas and temperate phage.....	74

## List of Tables

Table 2-1: Estimates of mutation effects and their interactions.....	16
Table 2-2: Primers used for phage mutagenesis .....	20
Table 2-3: Primers used for mutant verification .....	20
Table 3-1: Population doubling times of strains carrying RM systems.....	25
Table 3-2: Fractions of cells above the threshold in populations carrying RM systems.....	30
Table 3-3: Induction frequencies of SOS response inside microfluidic device. ....	35
Table 4-1: eop and eol of bacteria carrying different RM systems .....	56
Table 4-2: Model parameters.....	79
Table 4-3: Model initial conditions.....	80

## List of Symbols/Abbreviations

<b>CRISPR/Cas</b>	Clustered regularly interspaced short palindromic repeats, CRISPR-associated systems
<b>cfu</b>	Colony-forming units
<b>DDE</b>	Delayed differential equation
<b>eol</b>	Efficiency of lysogen formation
<b>eop</b>	Efficiency of plating/Efficiency of plaque formation
<b>M</b>	Methyltransferase/Modification enzyme
<b>MOI</b>	Multiplicity of infection
<b>pfu</b>	Plaque-forming units
<b>PDT</b>	Population doubling time
<b>R</b>	Restriction endonuclease/Restriction enzyme
<b>RM</b>	Restriction-modification
<b>wt</b>	Wild type



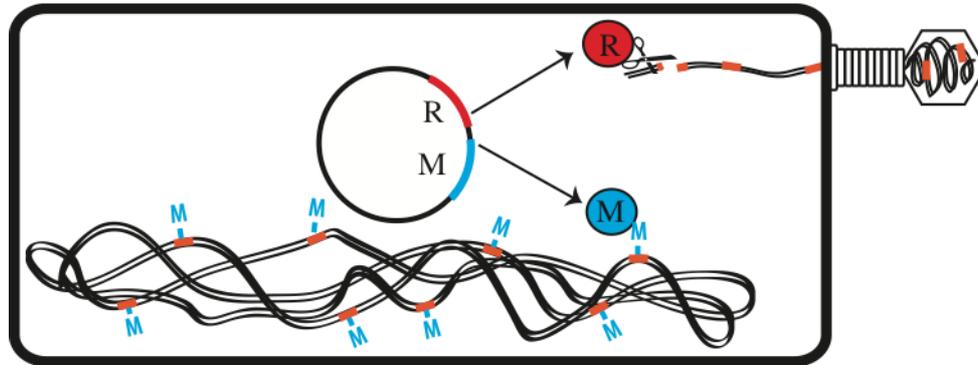
# 1 Introduction

Restriction-Modification (RM) systems are an integral part of the prokaryotic world. They are present in nearly all bacteria, as well as archaea, with multiple RM systems frequently coexisting in a single genome (Oliveira, Touchon, & Rocha, 2014). Although they were discovered more than six decades ago and their molecular mechanism of action has been studied in great detail, the important question regarding their role in natural ecosystems remains unanswered. Typically, RM systems are thought to represent a primitive form of innate immunity against infections by bacteriophages (phages) (Took & Dryden, 2005). Depending on the environment, phages can outnumber bacteria by 1-2 orders of magnitude (Wigington et al., 2016) and represent the major drivers of prokaryotic evolution (Koskella & Brockhurst, 2014). It comes as no surprise that bacteria would carry mechanisms to increase their chances of survival in the face of constant threat of phage attack. While it is beyond a doubt that RM systems can significantly lower the likelihood of phage infections, whether this is their evolutionary *raison d'être* is a different question, the answer to which is less clear and more difficult to find (Murray, 2002). My aim in this introductory chapter is to briefly summarize our understanding of the biology of these intriguing genetic elements and outline some of the open questions, which I think will need to be answered before we can fully appreciate the role that RM systems play in the ecology and evolution of bacteria and phages.

## 1.1 *Molecular mechanism and types of RM systems*

Most RM systems are composed of two components encoded by a pair of genetically linked genes. The first component is the restriction endonuclease (R), which recognizes well-defined DNA sequences called restriction sites and cleaves the DNA either within these sites, or in their proximity. The second component is the methyltransferase (M), also called the modification enzyme, which recognizes the same sequences as R and modifies them by covalently attaching a methyl group to one of the nucleotides (Arber & Linn, 1969). According to the traditional view, the role of R is to recognize and cleave heterologous unmodified DNA, such as that injected into the bacterium during phage infection. The role

of M is then to make sure that all bacterium's own restriction sites are methylated and thus protected from self-restriction (**Figure 1-1**).



**Figure 1-1: The mechanism of action of RM systems**

Methyltransferase (M) and restriction endonuclease (R) recognize the same specific sequences (in orange) called restriction sites. R recognizes and cleaves unmethylated restriction sites on the exogenous DNA and thus prevents phage infection. M methylates endogenous restriction sites and thus protects the host from self-restriction.

Despite their uniform mechanism of action, RM systems are a highly diverse group of genetic elements. They vary in gene arrangement, structural composition, type of sequence recognized, cleavage position and cofactor requirements. Based on these criteria they are classified into four main types (R. Roberts, 2003):

Unlike all other types, **Type I** RM systems are composed of three components. In addition to R and M, they contain a specificity subunit S. All three components form a single multi-subunit enzyme with both the endonuclease and methyltransferase activity. DNA cleavage is ATP dependent and takes place at a random distance from the recognition sequence. The first RM systems discovered (EcoKI, EcoAI, EcoBI) are type I (Loenen, Dryden, Raleigh, & Wilson, 2014).

**Type II** RM systems are the simplest, best studied, and most abundant type. In type II RM systems, R cleaves the DNA within or in close proximity to the recognition sequence independent of M and ATP (Loenen, Dryden, Raleigh, Wilson, & Murray, 2014). Based on their molecular properties, type II RM systems are divided into several sub-types. Type IIP RM systems code for functionally and structurally independent R and M enzymes, both of

which recognize identical palindromic sequences. The best-studied RM systems (EcoRI, EcoRII, EcoRV) are type IIP. R and M of type IIG RM systems, while functionally independent, are fused together into a single structural unit (Liang & Blumenthal, 2013). Although not well studied, Type IIG systems are very abundant in bacterial genomes (Oliveira et al., 2014).

**Type III** RM systems are structurally similar to type I, although they do not encode a separate specificity subunit. R and M of type III RM systems form multi-subunit enzymes, which act on two recognition sequences facing the opposite orientation. DNA cleavage is ATP-dependent. RM systems carried by phage P1 (EcoP1I, EcoP15I) are type III (Rao, Dryden, & Bheemanaik, 2014).

**Type IV** RM systems have been added to the portfolio only recently. They typically consist of only R, which recognizes and cleaves methylated sequences. McrBC from *Escherichia coli* is a well-studied example of a type IV RM (Loenen & Raleigh, 2014).

## **1.2 Abundance and mobility of RM systems in bacteria**

More than 10,000 RM systems have been found so far, with the number growing rapidly as more bacterial genomes are sequenced (R. Roberts, Vincze, Posfai, & Macelis, 2015). Naturally competent bacteria like *Helicobacter pylori*, *Haemophilus influenzae*, *Neisseria gonorrhoeae*, *Neisseria meningitidis* and *Methanococcus jannaschii* typically carry a large number of different RM systems. On the other hand, no RM system has yet been described in obligatory intracellular parasites such as *Rickettsia*, *Chlamydia* and *Coxiella* (Blumenthal & Cheng, 2002). The reasons for such enormous variability in RM systems and the factors that determine the optimal number of RM systems per genome are not understood.

RM systems are highly mobile genetic elements. They often reside on plasmids, phages, transposons or integrons, suggesting their tendency to undergo horizontal gene transfer (Furuta & Kobayashi, 2012). Homologous RM systems were found in distantly related organisms such as *Methanobacterium thermoformicum* and *N. gonorrhoeae* (Nölling & Vos, 1992) and intra-genomic movement of an IS-linked RM system has been directly observed in laboratory experiments (Takahashi, Ohashi, Sadykov, Mizutani-Ui, & Kobayashi, 2011). However, the evolutionary relationships between different RM systems are typically

difficult to disentangle (Blumenthal & Cheng, 2002). Sequences of different RM systems are typically very diverse and indicate that RM systems have evolved multiple times independently (Wilson & Murray, 1991). Since cognate methyltransferases and endonucleases recognize the same nucleotide sequences, one would expect at least the target recognition domains of cognate R and M pairs to be homologous. Interestingly, this is not the case (Chandrasegaran & Smith, 1988; Wilson & Murray, 1991). Moreover, methyltransferases are typically more conserved than endonucleases, whose sequences are highly divergent (Pingoud et al., 2002). There is currently no satisfactory explanation for the seemingly different rates of evolution of the two enzymes.

### **1.3 Possible roles of RM systems**

RM systems were originally discovered due to their ability to protect bacteria from phage infections (Bertani & Weigle, 1953) and they were subsequently labeled as a type of prokaryotic immunity. However, there has so far been only one attempt to directly test whether RM systems could have evolved and be maintained in nature due to their ability to protect bacteria from phages (Korona & Levin, 1993). In these experiments, the authors directly competed *E. coli* strains with and without an RM system in the presence of three different phage species, and observed that carrying an RM system was of little benefit to the bacteria. Instead of relying on the RM system as a general mechanism of defense, both RM+ and RM- populations evolved three distinct phage receptor mutations, each to protect it from a specific phage (Korona & Levin, 1993). Rather than rejecting the hypothesis according to which RM systems evolved and are maintained as a resistance mechanism, this intriguing result opened up a number of new questions. To what extent is this result dependent on the environment and to what extent can it be generalized to other RM systems and other phages? What are the advantages of receptor resistance versus RM-based immunity? Why have bacteria evolved various mechanisms of immunity if receptor resistance is a readily available and efficient method of phage resistance? Unfortunately, these questions have received little attention and whether or not RM systems could have evolved as an immune system is still unknown.

An alternative hypothesis for the potential role of RM systems states that, instead of

protecting bacteria from parasites, RM systems themselves are parasitic pieces of “selfish” DNA. According to this hypothesis, RM systems confer no benefit to the host bacterium and persist solely because, once acquired, they cannot be lost without death of the former host bacterium (Naito, Kusano, & Kobayashi, 1995). Such “genetic addiction” behavior has been demonstrated in laboratory conditions for several type II RM systems (Kobayashi, 2001). In these experiments, induced loss of a plasmid carrying an RM system resulted in post-segregational killing, similar to that caused by toxin-antitoxin (TA) systems (Hayes, 2003), albeit caused by a different mechanism (Handa & Kobayashi, 1999). It has to be noted though that post-segregational killing can only operate in the context of type II RM systems and that type I, III and IV RM systems do not and cannot, based on their mechanism of action, behave selfishly (O’Neill, Chen, & Murray, 1997). Moreover, similarly to the “phage immunity” hypothesis, it has not been convincingly shown that RM systems can increase their frequency in a population solely due to their selfish behavior. Finally, while post-segregational killing is an efficient mechanism of preventing loss of RM systems *en bloc*, costly RM systems could still be lost in two steps if R is lost first and M second. Selfishness of RM systems alone is thus unlikely to fully explain the high abundance of RM systems.

Several alternative hypotheses for the possible evolutionary roles of RM systems have been proposed and recently reviewed in detail (Vasu & Nagaraja, 2013). In addition to the above mentioned, RM systems could regulate gene flux (Oliveira, Touchon, & Rocha, 2016), maintain species identity (Jeltsch, 2003), promote recombination (Chang & Cohen, 1977) and stabilize genomic islands (Kusano, Naito, Handa, & Kobayashi, 1995). It is important to note that these hypotheses are not mutually exclusive and it is possible that the widespread occurrence of RM system is caused by more than one factor.

#### **1.4 RM systems as a stochastic molecular mechanism**

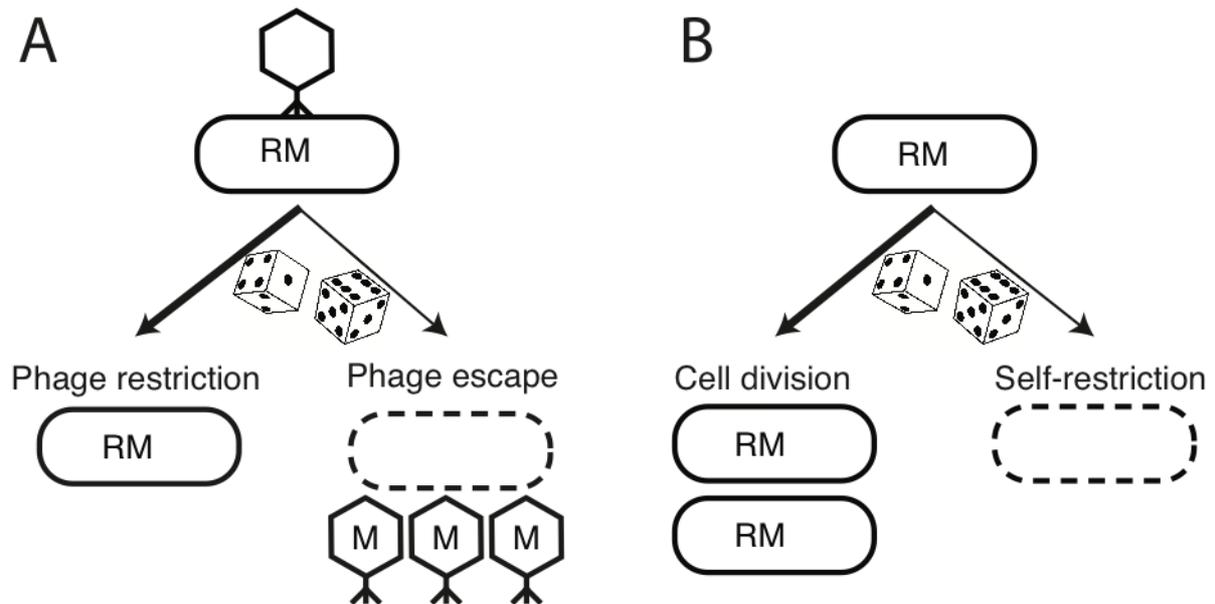
In the very early experiments that lead to the discovery of RM systems (Bertani & Weigle, 1953; Luria & Human, 1952), it has been observed that although RM systems significantly reduce the likelihood of phage infections, they are not impermeable and that there is a non-zero probability that an infecting phage escapes restriction and produces methylated progeny. In other words, while the majority of genetically identical phages infecting

genetically identical bacteria in an identical environment gets cleaved and the infection aborted, in a small subpopulation of infected bacteria, phages will get methylated instead of restricted, seemingly at random. RM systems thus represent an intrinsically stochastic molecular system. It is intriguing that stochasticity as a property of RM systems has over decades of research been overlooked, and its causes and consequences are unknown. Moreover, the importance of understanding the consequences of stochasticity on ecological and evolutionary dynamics goes beyond the context of RM systems. Recently, a great deal of scientific interest has been focused on the effects of “molecular noise” (McAdams & Arkin, 1999) at the level of single cells and how it can lead to heterogeneity in clonal microbial populations (Elowitz, Levine, Siggia, & Swain, 2002). Many mechanisms capable of generating heterogeneity were identified (Avery, 2006), but the question of to what extent do stochastic events affect dynamics of populations has remained unanswered.

Is it possible that random events occurring at the level of individuals could affect the ecological and evolutionary dynamics of bacteria and phages and leave a detectable footprint in their genomes? It is known that restriction sites of many RM systems are significantly underrepresented in phage genomes (Karlin, Burge, & Campbell, 1992; Rocha, Danchin, & Viari, 2001). This phenomenon, also called restriction site avoidance, is often thought to result from selection for phages with fewer restriction sites due to their increased probability of escape and this explanation is used as an argument in support of the hypothesis that the primary role of RM systems is to prevent phage infections (Tock & Dryden, 2005). However, at present we do not understand why some phages escape restriction while others get cleaved, nor do we know how the probability of escape depends on the number of phage restriction sites. Any connections between phage escape as a stochastic process and restriction site avoidance in phage genomes are thus difficult to establish.

Interestingly, it has been shown that bacteria avoid using restriction sites in their genomes to a similar, or greater, extent than phages (Rocha et al., 2001). The causes of restriction site avoidance in bacterial genomes are unknown. It has been hypothesized that, as a consequence of the stochastic nature of the underlying molecular mechanism, RM systems occasionally attack chromosomal DNA of their own hosts and that this leads to natural selection for bacteria with fewer restriction sites due to their reduced likelihood of self-

restriction. This can indeed be a plausible explanation since the concentration of R and M enzymes could be subject to random fluctuations either as a result of stochastic gene expression (Elowitz et al., 2002), or random partitioning following cell division (Huh & Paulsson, 2011). However, self-restriction as a form of autoimmunity caused by RM systems has not yet been experimentally observed and whether it occurs in bacterial populations is not known.



**Figure 1-2: Effects of stochasticity on the function of RM systems**

**(A)** A bacterium carrying an RM system infected by a phage can either successfully restrict the phage, resulting in host survival (high probability event), or it can fail to restrict the phage, leading to phage escape, production of methylated progeny phages and host death (low probability event).

**(B)** A bacterium carrying an RM system can either grow and divide normally (high probability event), or get attacked by the resident RM system, leading to self-restriction and host death (low probability event).

### 1.5 Other mechanisms of Prokaryotic immunity

Regardless of their true evolutionary role, RM systems can clearly lower the likelihood of phage infections. In addition to RM, there are multiple other mechanisms that provide bacteria with protection from phages by interfering with various stages of the phage life cycle (Labrie, Samson, & Moineau, 2010). One of the most intriguing questions regarding the interactions between bacteria and phages is why such variety of resistance mechanisms

evolved, why it is maintained and what mechanisms are optimal under what environmental conditions (Houte, Buckling, & Westra, 2016).

Most phages infect bacteria through specific receptors located at the surface of the host bacterium. For example, the classic phage  $\lambda$  interacts with the maltose transporter LamB. Laboratory experiments in which bacteria grow in presence of phages typically lead to appearance and increase in frequency of **receptor resistance** mutants. Such mutants can prevent phage adsorption either by lowering the expression, or complete inactivation of the phage receptor. Loss of receptors is often associated with significant fitness costs in the absence of phage, which leads to a tradeoff between growth and resistance (Westra et al., 2015). In addition to mutations directly affecting expression or structure of phage receptors, bacteria can respond to phages by overproduction of extracellular polysaccharide, which masks the receptors, lowers the phage adsorption and causes the typical mucoid phenotype (Ohshima, Schumacher-Perdreau, Peters, & Pulverer, 1988).

Similarly to RM systems, genomes of many bacteria and archaea carry systems composed of clustered, regularly interspaced, short palindromic repeats (CRISPR), together with the CRISPR-associated (Cas) proteins. **CRISPR/Cas** systems provide bacteria and archaea with a mechanism of adaptive immunity to phages (Barrangou et al., 2007). In contrast to RM systems, which cleave phage DNA at defined sequences regardless of whether they have been previously exposed to the targeted phage, CRISPR/Cas can acquire sequences from infecting phage and use these sequences (referred to as protospacers) for recognition and specific cleavage of phage genomes. CRISPR/Cas are a highly diverse class of genetic elements and are divided into a number of types and subtypes (Makarova et al., 2015). Furthermore, just like RM systems, CRISPR/Cas are useful tools of genetic engineering and while a lot of research has been devoted to the function of CRISPR/Cas and their applications, the role they play in the ecology and evolution of bacteria and phages has received significantly less attention.

**Abortive infection** systems provide bacteria with protection from phages by limiting their spread in bacterial populations (Chopin, Chopin, & Bidnenko, 2005). When a bacterium carrying an abortive infection system is infected by a phage sensitive to its effects, the system will trigger a response in the form of a programmed cell death. As a result, the

infecting phage will not complete its life cycle and spread of infection will be halted. Interestingly, abortive infection systems are frequently carried by prophages (temperate phages persisting in bacteria as parts of their genome) (Houte, Buckling, et al., 2016). The best studied abortive infection system is the *rII* exclusion system carried by phage  $\lambda$  (Parma et al., 1992).

**Superinfection exclusion** systems, also frequently carried by prophages, provide bacteria with immunity to co-infecting phages by directly interfering with phage replication via mechanisms other than programmed cell death (Bondy-Denomy et al., 2016). Superinfection exclusion allows prophages to prevent secondary infections by phages of the same or different type. Perhaps the best-studied example of a superinfection exclusion system is the one mediated by the  $\lambda$  CI repressor, which prevents replication of  $\lambda$  phages infecting  $\lambda$  lysogens (Fogg, Allison, Saunders, & McCarthy, 2010).

## **1.6 Questions addressed in this thesis**

In this thesis, I adopt the framework of RM systems functioning as a mechanism of prokaryotic immunity and explore its implications in various scenarios. In all experiments, I use two model RM systems originally isolated from *E. coli*, EcoRI and EcoRV, as well as the classic phage  $\lambda$ . While, the two RM systems and the phage are very well studied at the molecular level, we know very little about the nature and dynamics of their interactions. Throughout the thesis, I focus on stochasticity as a basic property of RM systems and ask how it affects interactions between RM systems, bacteria and phages.

In Chapter 2, I study the process of phage escape and its implications for evolution of restriction site avoidance. The chapter addresses the following set of questions: Do mutations in restriction sites increase the likelihood of phage escape? What are the underlying molecular events that lead to phage escape? What factors determine the probability of phage escape as a key parameter of interactions between RM systems and phages?

In Chapter 3, I study interactions between RM systems and bacteria. The main questions addressed are: Why are restriction sites underrepresented in bacterial genomes? Do RM

system exert a fitness cost on their host and if so, is the cost related to function of RM systems as a mechanism of phage resistance? How do bacteria deal with the inherent stochasticity of RM systems?

In Chapter 4 I explore the nature and dynamics of interactions between RM systems, bacteria and phages all together. Specifically, I focus on interactions between bacteria carrying RM systems and temperate phages, which can both kill the infected bacteria, as well as provide them with potentially beneficial genes. In this chapter, I ask the following questions: Are RM systems an efficient mechanism of immunity to temperate phages? Does protection from lethal infections necessarily reduce the probability of acquisition of potentially beneficial prophages? Can RM systems distinguish between lytic and lysogenic infections?

## 2 Effects of mutations in phage restriction sites during escape from restriction-modification

This chapter was originally published in: Pleška M and Guet CC (2017) Effects of mutations in phage restriction sites during escape from restriction–modification. *Biology Letters*. 13, 20170646.

### 2.1 Summary

Restriction-modification systems are widespread genetic elements that protect bacteria from bacteriophage infections by recognizing and cleaving heterologous DNA at short, well-defined sequences called restriction sites. Bioinformatic evidence shows that restriction sites are significantly underrepresented in bacteriophage genomes, presumably because bacteriophages with fewer restriction sites are more likely to escape cleavage by restriction-modification systems. However, how mutations in restriction sites affect the likelihood of bacteriophage escape is unknown. Using the bacteriophage  $\lambda$  and the restriction-modification system EcoRI, we show that while mutation effects at different restriction sites are unequal, they are independent. As a result, the probability of bacteriophage escape increases with each mutated restriction site. Our results provide direct experimental support for restriction site avoidance as an effective response to selection imposed by restriction-modification systems and offer an insight into the events underlying the process of bacteriophage escape.

## 2.2 Introduction

Bacterial viruses, also called bacteriophages (phages) are the most abundant biological entities on Earth and as such, they represent a major driving force of bacterial evolution (Koskella & Brockhurst, 2014). While temperate phages can, with a small probability, enter genomes of their hosts and potentially contribute genes that increase bacterial fitness (Canchaya, Fournous, & Brüssow, 2004), the vast majority of infections by both temperate and virulent phages are lethal for infected bacteria. To protect themselves, many bacteria utilize a wide variety of phage resistance mechanisms a large group which are based on recognizing and destroying heterologous phage DNA (Houte, Buckling, et al., 2016). Restriction-modification (RM) systems represent the first discovered (Luria & Human, 1952), the simplest and one of the most prevalent (Oliveira et al., 2014) of such mechanisms.

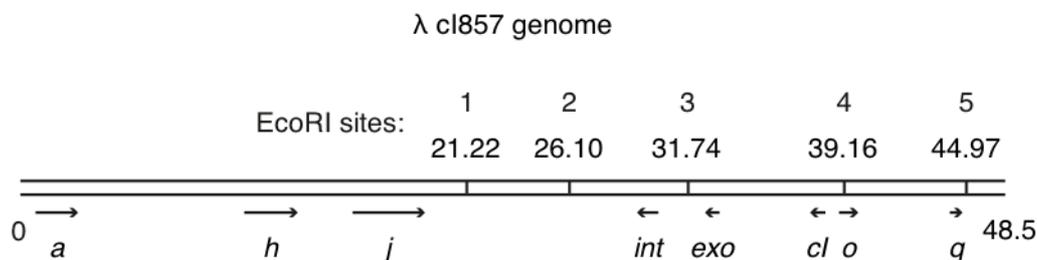
Most RM systems are composed of two enzymatic activities: the restriction activity of a restriction endonuclease (R) and the modification activity of a methyltransferase (M). Both R and M typically recognize and act on well-defined, short (4-8 bp) DNA sequences termed restriction sites. Upon infection, R recognizes the restriction sites on the phage DNA as non-self and cleaves it, thus aborting the infection. However, there is a non-zero probability (typically at the order of  $10^{-5}$ ) that instead of restricted, the phage restriction sites will be erroneously modified by M, whose primary role is to methylate restriction sites contained in the bacterium's own DNA and prevent self-restriction (Bickle & Krüger, 1993). What causes a fraction of phages to escape restriction is not understood, as are not the factors determining the size of this fraction itself.

Interestingly, Genomes of many phages display a significant underrepresentation of restriction sites (Karlin et al., 1992; Krüger & Bickle, 1983; Rocha et al., 2001). This underrepresentation, also termed restriction site avoidance, is thought to result from natural selection favoring phages with mutations in restriction sites due to their increased probability of escape. In our work, we study the relationship between mutations in restriction sites and the probability of phage escape.

## 2.3 Results

### 2.3.1 Effects of mutations in restriction sites are unequal

To study the effect of mutations in phage restriction sites on the probability of phage escape, we first constructed five  $\lambda$  *cl857* mutants, each with a single point mutation in one of the five EcoRI restriction sites (**Figure 2-1**). The probability of phage escape, measured as the efficiency of plating (*eop*) (Material and Methods), of all five mutants was considerably higher than the probability of escape of the wild type  $\lambda$  *cl857* (**Figure 2-2A**). Each mutation thus increased the likelihood of phage escape. While mutations at sites 1, 3, 4, and 5 all increased *eop* by approximately an order of magnitude, the effect of mutation at site 2 was considerably smaller. This result was in accord with previous studies, showing that this particular EcoRI restriction site is cleaved with a lower efficiency both *in vitro* (Berkner & Folk, 1983) and *in vivo* (Murray & Murray, 1974). The lower likelihood of cleavage is possibly a result of the significantly reduced GC-content in an approximately five-kb-long region, in which the restriction site is located (Berkner & Folk, 1983).



**Figure 2-1: Genetic map of  $\lambda$  *cl857***

Locations (in kb) of five restriction sites are specified. Locations of characteristic  $\lambda$  genes are shown for reference.

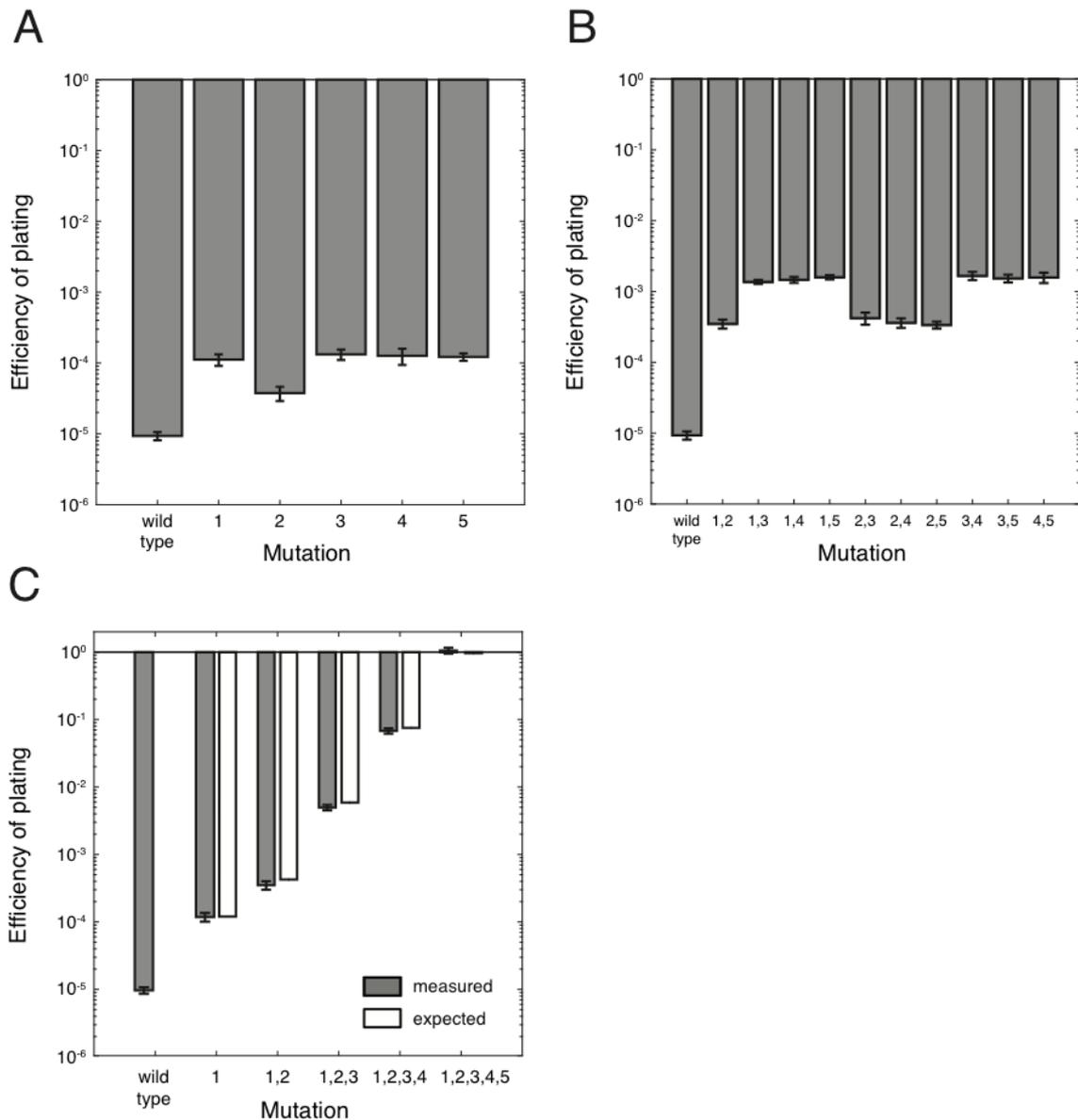
### 2.3.2 Effects of mutations in phage restriction sites are independent

We next asked whether effects of mutations in restriction sites change when they occur in combinations, for example as a result of relative positioning of individual restriction sites (Rau & Sidorova, 2010). To this end, we constructed a set of ten  $\lambda$  *cl857* mutants, each with a different pairwise combination of mutations. The *eop* of all ten double mutants was higher than the wild type, although apparent differences among mutants were observed (**Figure**

**2-2B**). Namely, the *eop* of phages including a mutation at site 2 was lower as compared to other mutants. We tested whether interactions alter the effect of mutations occurring in combinations by fitting a linear regression model with interaction terms to data obtained for both single and double mutants. The effects of all five individual mutations were highly significant, with the effect at site 2 being the smallest (**Table 2-1**). In contrast, all interaction terms were below the significance threshold ( $\alpha=0.05$ ), implying that mutation effects are independent.

### **2.3.3 Effects of mutations in restriction sites are multiplicative**

To elucidate the dependence the phage escape probability on the number of restriction sites, we created a third set of mutants by introducing mutations consecutively. As shown in **Figure 2-2C**, each additional mutated restriction site considerably increased the *eop*. The mutant carrying all five mutations formed plaques on bacteria carrying EcoRI with the same probability as on bacteria devoid of the RM system (*eop*=1). This result indicated that point mutations in restriction sites were sufficient to completely abolish cleavage. We compared the measured *eop* values to the expected values calculated based on individual mutation effects estimated in **Table 2-1**, assuming complete independence of mutation effects. The measured and the expected *eop* values were in a good agreement, further supporting the independence of mutation effects in phage restriction sites.



**Figure 2-2: Effect of mutations in restriction sites on efficiency of plating**

**(A)** Efficiencies of plating of wild type  $\lambda$  *cI857* and five single mutants. The numbering of restriction sites corresponds to that shown in **Figure 2-1**.

**(B)** Efficiencies of plating of wild type  $\lambda$  *cI857* and all double mutants.

**(C)** Experimentally observed efficiencies of plating of the wild type  $\lambda$  *cI857* and mutants with consecutively added mutations are shown as grey bars. White bars represent expected efficiencies of plating calculated based on effects of individual mutations, assuming complete independence of mutation effects. In all panels, means of six independent biological replicates from three separate experiments are shown except for wild type, which shows 18 independent biological replicates from three separate experiments. Error bars represent 95% confidence intervals.

**Table 2-1: Estimates of mutation effects and their interactions**

	Est. effect	St. Error	p-Value
<b>Main effects:</b>			
Intercept (wt <i>eop</i> )	9.45E-06	1.05	<0.001
RS1	12.43	1.11	<0.001
RS2	3.56	1.11	<0.001
RS3	13.87	1.11	<0.001
RS4	12.81	1.11	<0.001
RS5	12.81	1.11	<0.001
<b>Interaction effects:</b>			
RS1:RS2	0.84	1.16	0.234
RS1:RS3	0.84	1.16	0.27
RS1:RS4	0.97	1.16	0.856
RS1:RS5	1.06	1.16	0.704
RS2:RS3	0.90	1.16	0.455
RS2:RS4	0.84	1.16	0.235
RS2:RS5	0.78	1.16	0.108
RS3:RS4	0.99	1.16	0.945
RS3:RS5	0.91	1.16	0.541
RS4:RS5	1.00	1.16	0.983

## 2.4 Discussion

Our results demonstrate that point mutations in phage restriction sites can substantially increase the probability of phages escaping restriction, thus providing direct experimental support for restriction site avoidance as an adaptive response to selection imposed by RM systems. Phages with complete avoidance of restriction sites are frequently found in nature (Korona, Korona, & Levin, 1993). Some of the most avoided sequences are recognition sequences of type IIP RM systems, such as EcoRI studied here, which recognize and cleave DNA at individual restriction sites. However, more complex patterns of selection can be observed in the case of RM systems requiring two recognition sequences for cleavage. For example, EcoRII, a type IIE RM system, cleaves the DNA only if it recognizes two proximate recognition sequences (Krüger, Kupper, Meisel, Reuter, & Schroeder, 1995), whereas EcoP1I, a type III RM system, cleaves the DNA upon recognition of two opposing asymmetric recognition sequences (Meisel, Bickle, Kriiger, & Schroeder, 1992). Accordingly, in the genome of phage T7, recognition sequences of EcoRII are distantly apart, whereas those of EcoP1 are all facing the same direction (Tock & Dryden, 2005).

In addition to evolution of restriction site avoidance, our results offer an insight into the molecular events underlying phage escape as a probabilistic process. At the level of single bacteria, reaction events are often stochastic as a result of various mechanisms including “noise” in gene expression (McAdams & Arkin, 1997) and stochastic enzymatic kinetics in general. Despite being known for decades, the molecular events underlying phage escape as a stochastic reaction event are not well understood. Our results allow us to distinguish between two hypotheses: If the main factor responsible for a fraction of phages escaping restriction was preexisting phenotypic variability in a bacterial population, the probability of phage escape should reflect the size of the susceptible subpopulation and thus be relatively independent of the number of phage restriction sites. On the other hand, if the main factor responsible for mixed outcomes was competition between R and M in recognizing the newly appearing restriction sites, the probability of phage escape should be equal to the product of probabilities with each restriction sites gets methylated instead of cleaved. Our results are in good agreement with the latter hypothesis and thus emphasize the importance of molecular noise in the biology of RM systems (Pleška et al., 2016).

## **2.5 Material and Methods**

### **2.5.1 Experimental system**

As a model system, we used the classic RM system EcoRI and bacteriophage  $\lambda$  variant *ci857*, which carries five EcoRI restriction sites (GAATTC) (**Figure 2-1**).  $\lambda$  *ci857* is a temperature-sensitive mutant, which behaves as an obligatory lytic phage at temperatures above 30°C (Oppenheim & Salomon, 1972). All our experiments were performed at 37°C. *Escherichia coli* strain MG1655 was used in all experiments. The EcoRI RM system was carried on a plasmid pBR322 $\Delta$ Ptet EcoRI (R+M+) (Pleška et al., 2016). As a RM- reference, we used a strain carrying the pBR322 $\Delta$ Ptet plasmid.

### **2.5.2 Culture conditions**

To provide a constant and well reproducible number of bacteria in a comparable physiological state, all cultures used to measure *eop* were grown from fresh colonies for 24 hours in M9 medium with maltose (1x M9 salts (12.8 g/l Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O, 3 g/l KH<sub>2</sub>PO<sub>4</sub>, 0.5 g/l NaCl, 1 g/l NH<sub>4</sub>Cl), 0.4% maltose, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>) to increase phage adsorption. Ampicillin (100 µg/ml) was added to the media to select for plasmid maintenance. Phage plates (1% tryptone, 0.1% yeast extract, 0.8% NaCl, 1% agar, 0.01% glucose, 0.2 mM CaCl<sub>2</sub>) and Phage soft agar (the same as above, but with 0.7% agar) were used for plaque enumeration.

### **2.5.3 Phage mutagenesis and lysate preparation**

All phage lysates were prepared by plate lysis. Briefly, individual plaques were picked and resuspended in 3 ml of phage soft agar together with 100 µl of overnight *E. coli* MG1655 cultures grown in M9 medium with maltose. After 8 hours, the soft agar was scraped and resuspended in 10 ml of SM buffer, centrifuged, filtered (0.2 µm) and stored at 4°C. All phage mutants originate from a single  $\lambda$  *ci857* plaque and were constructed by recombineering as described in (Oppenheim, Rattray, Bubunenko, Thomason, & Court, 2004). The oligonucleotides used for mutagenesis are listed in **Table 2-2**. The

oligonucleotides were designed such that a single point mutation is introduced into the EcoRI restriction site. For those EcoRI sites, which are located inside a coding sequence (1-4), the mutations introduced were picked such that the amino acid sequence of the gene product is preserved. Mutants were identified using a PCR method: 16 individual plaques were picked and resuspended in 50  $\mu$ l of water. 2  $\mu$ l were then used as a template in a PCR reaction together with primers designed to amplify 500 bp upstream and 500 bp downstream of the restriction site into which the mutation was introduced. OneTaq<sup>®</sup> Quick-Load<sup>®</sup> 2X Master Mix with Standard Buffer (NEB) was used for all reactions according to the standard protocol. The primers used for verification are listed in **Table 2-3**. Successful recombinants were identified by mixing 4  $\mu$ l of the PCR product with EcoRI-HF<sup>®</sup> in a standard digestion reaction (10  $\mu$ l total volume). Isolates that gave PCR products which were not digested were identified as recombinants and were plated first on a lawn of *E. coli* MG1655, from which individual plaques were picked and used to prepare working lysates. All phage lysates were stored and diluted in SM buffer (100 mM NaCl, 8 mM MgSO<sub>4</sub>, 200 mM Tris-Cl (pH 7.5)). The titer of all lysates used was adjusted to approximately  $2 \cdot 10^8$  pfu/ml.

**Table 2-2: Primers used for phage mutagenesis**

<b>Name</b>	<b>Sequence<sup>a</sup></b>
Lam_RI(21,226)_ko	CCGTTGCAGATGTTCTTGAATACCTTGGGGCCGGTGAG <u><b>AACTC</b></u> GGCCTTTCCGGCAGGTGCGCCGATCCCGTGGCCATCA
Lam_RI(26,104)_ko	CAGCAATAGTTTAAAATCACTAGGCGATCTCCGCTTAG <u><b>AGTTCA</b></u> TTTCAGCATTTATTGGTTGTATGAGAGTAGATAGAA
Lam_RI(31,747)_ko	GGGAAAACAGTACGAGAACGACGCCAGAACCCTGTTT <u><b>GAGTTC</b></u> ACTTCCGGCGTGAATGTTACTGAATCCCCGATCATCT
Lam_RI(39,168)_ko	CTATTACAAAAGAAAAAAGAAAAGATTATTCGTCAGAG <u><b>AACTC</b></u> TGGCGAATCCTCTGACCAGCCAGAAAACGACCTTTCT
Lam_RI(44,972)_ko	GCACAACCCAACTGAGCCGTAGCCACTGTCTGTCCT <u><b>GAACTCA</b></u> TTAGTAATAGTTACGCTGCGGCCTTTTACACATGAC

<sup>a</sup> Restriction site sequences are shown in bold, mutations are underlined

**Table 2-3: Primers used for mutant verification**

<b>Name</b>	<b>Sequence</b>
fw_Lam_RI(21,226)	AAAGGGGATAGTGCAGCTCA
rv_Lam_RI(21,226)	CAATACCCTGTGTGCTGGTT
fw_Lam_RI(26,104)	TCAATATCCGGACGGATAAT
rv_Lam_RI(26,104)	TTGAAAATGAAAGCGTCCTT
fw_Lam_RI(31,747)	ATTTCCGGATAACAGAAAGGC
rv_Lam_RI(31,747)	GCATACACTGCAGAACGTCA
fw_Lam_RI(39,168)	CCAGATGGAGTTCTGAGGTC
rv_Lam_RI(39,168)	TTTTCGTCGTAAGTTCCGG
fw_Lam_RI(44,972)	TCGCAGACAACATTTTGAAT
rv_Lam_RI(44,972)	AGCAGCGAAGCGTTTGATA

#### 2.5.4 Measurements of phage escape probability

As a measure of the probability of escape, we used the efficiency of plating defined as  $eop = \frac{pfu_{RM}}{pfu_{total}}$ , where  $pfu_{RM}$  is the number of plaque forming units ( $pfu$ ) obtained on lawns of bacteria carrying the EcoRI RM system and  $pfu_{total}$  is the total number of  $pfu$  obtained on the reference RM- strain. For each measurement, 10  $\mu$ l of serially diluted lysates was mixed with 0.1 ml of bacterial cultures in 3 ml of soft agar and spread on phage plates such that 100-300  $pfu$  were obtained on each plate. In each measurement,  $pfu_{total}$  was at least an order of magnitude lower than the total number of bacteria plated ( $\approx 10^8$ ) so that the vast majority of infections correspond to a single phage infecting a single bacterium. Method of preparation of lysates and bacterial cultures is described in the Supplementary material and Methods.

#### 2.5.5 Statistical analysis

The effects of mutations and their interactions were calculated by fitting a single multivariate linear regression model with interaction terms to the data shown in **Figure 2-2A** and **Figure 2-2B**, with  $\log(eop)$  as the dependent continuous variable and presence/absence of each restriction site as categorical independent variables. A single model was fit to the data obtained from experiments with both individual mutations and their pairwise combinations. Normal distribution of errors was verified by residual analysis.

## 3 Bacterial autoimmunity due to a restriction-modification system

This chapter was originally published in: Pleška M, Qian L, Okura R, Bergmiller T, Wakamoto Y, Kussell E, & Guet CC (2016). Bacterial autoimmunity due to a restriction-modification system. *Current Biology*, 26(3), 404-409. The chapter was written in collaboration with Long Qian and Edo Kussell (New York University, NY, USA), who performed the bioinformatics analysis, and Reiko Okura and Yuichi Wakamoto (University of Tokyo, Tokyo, Japan), who performed and analyzed the microfluidic experiments.

### 3.1 Summary

Restriction-modification (RM) systems represent a minimal and ubiquitous biological system of self/non-self discrimination in prokaryotes (Oliveira et al., 2014), which protects hosts from exogenous DNA (Murray, 2002). The mechanism is based on the balance between methyltransferase (M) and cognate restriction endonuclease (R). M tags endogenous DNA as self by methylating short specific DNA sequences called restriction sites, while R recognizes unmethylated restriction sites as non-self and introduces a double-strand DNA (dsDNA) break (Arber & Dussoix, 1962). Restriction sites are significantly underrepresented in prokaryotic genomes (Elhai, 2001; Gelfand & Koonin, 1997; Karlin et al., 1992; Rocha et al., 2001), suggesting that the discrimination mechanism is imperfect and occasionally leads to autoimmunity due to self DNA cleavage (self-restriction) (Qian & Kussell, 2012). Furthermore, RM systems can promote DNA recombination (Chang & Cohen, 1977) and contribute to genetic variation in microbial populations, thus facilitating adaptive evolution (Asakura, Kojima, & Kobayashi, 2011). However, cleavage of self DNA by RM systems as elements shaping prokaryotic genomes has not been directly detected and its cause, frequency, and outcome are unknown. We quantify self-restriction caused by two RM systems of *Escherichia coli* and find that, in agreement with levels of restriction site avoidance, EcoRI, but not EcoRV, cleaves self DNA at a measurable rate. Self-restriction is a stochastic process, which temporarily induces SOS response, and is followed by DNA repair,

maintaining cell viability. We find that RM systems with higher restriction efficiency against bacteriophage infections exhibit a higher rate of self-restriction, and that this rate can be further increased by stochastic imbalance between R and M. Our results identify molecular noise in RM systems as a factor shaping prokaryotic genomes.

## 3.2 Results

### 3.2.1 EcoRI, but not EcoRV, induces DNA damage in host bacteria

We hypothesized that natively occurring RM systems cause occasional self-restriction and that this is detrimental to their host bacteria. To test this hypothesis, we compared population doubling times of *E. coli* MG1655 (wild-type) with plasmids carrying EcoRI or EcoRV RM systems (R+M+) expressed from their native promoters, respective control plasmids deficient in R activity (R-M+), and the plasmid backbone control (R-M-) (**Figure 3-1A**). Population doubling times of cells carrying EcoRI and EcoRV (R+M+) plasmids did not significantly differ from the controls (**Table 3-1**), indicating that self-restriction is either rare, and/or its effect is small due to the ability of wild-type cells to repair DNA damage (Cromie & Leach, 2001; Heitman, Ivanenko, & Kiss, 1999). We observed no measurable fitness effect in direct competitions between (R+M+) vs. (R-M+) strains of EcoRI and EcoRV in rich medium (M9, 0.4% glucose, 0.2% casamino acids)(**Figure 3-1B**), but observed decreased fitness due to EcoRI (R+M+) in minimal medium (M9, 0.4% glucose) (**Figure 3-1C**). Earlier studies have shown that induced chronic dsDNA breaks occurring once per replication cycle have only a small effect (0.6%) on the proliferation rate of wild-type *E. coli* (Darmon, Eykelenboom, Lopez-Vernaza, White, & Leach, 2014), and that the capacity to repair DNA damage is limited by resource availability (Sargentini, Diver, & Smith, 1983).

To test whether DNA damage occurs at elevated levels in populations carrying EcoRI, we measured the population doubling time of the *recA* knockout ( $\Delta recA$ ) strain carrying the RM plasmids. RecA is an essential component of DNA repair and, unlike wild-type cells, *recA* mutants are sensitive to self-restriction provoked by artificially induced imbalance between R and M expression (**Figure 3-2A**). Deleting *recA* increased the population doubling time of all strains by approximately six minutes (15% of the wild-type doubling time) (**Table 3-1**), reflecting the inability of  $\Delta recA$  cells to repair spontaneous DNA damage (Pennington & Rosenberg, 2007). Presence of the plasmid expressing EcoRI (R+M+) increased the doubling time of  $\Delta recA$  cells significantly by an additional three minutes as compared to the EcoRI (R-M+) control. In contrast, the EcoRV (R+M+) plasmid had no statistically significant effect on growth of the  $\Delta recA$  strain. The results suggested that EcoRI increased the amount of DNA

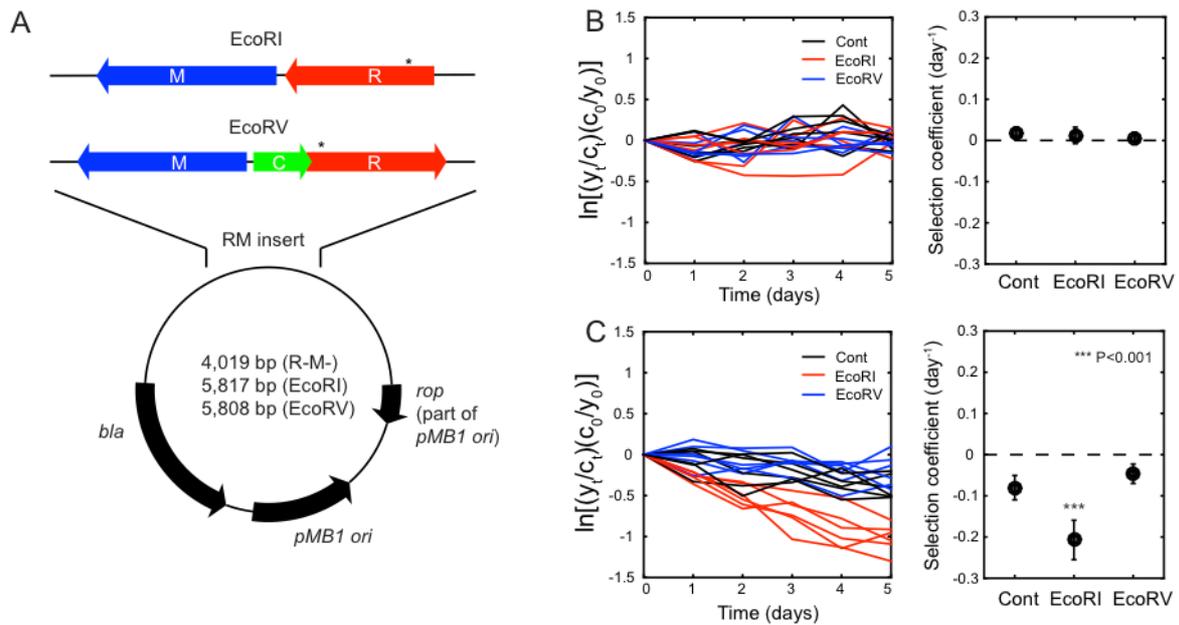
damage in the population and that RecA alleviated most of the negative effect on growth. The mechanism of restriction alleviation, which prevents self-restriction by type I RM systems, is unlikely to affect our estimates, since both EcoRI and EcoRV are type II RM systems and therefore insensitive to restriction alleviation (Makovets, Powell, Titheradge, Blakely, & Murray, 2004).

**Table 3-1: Population doubling times of strains carrying RM systems**

Host	Plasmid	PDT (min) <sup>a</sup>	SD (min) <sup>a</sup>	P-value <sup>a, b</sup>
wild-type	Control (R-M-)	40.09	1.49	
MG1655	EcoRI (R+M+)	40.78	0.68	0.429
	EcoRI (R-M+)	40.68	0.84	0.492
	EcoRV (R+M+)	40.79	0.57	0.422
	EcoRV (R-M+)	40.19	1.19	0.908
<i>ΔrecA</i>	Control (R-M-)	46.06	0.59	
	EcoRI (R+M+)	48.98	0.75	0.003
	EcoRI (R-M+)	46.20	0.99	0.854
	EcoRV (R+M+)	46.84	1.14	0.327
	EcoRV (R-M+)	45.67	1.08	0.626

<sup>a</sup> Calculated from three independent experiments, each with six biological replicates. PDT - population doubling time, SD - standard deviation.

<sup>b</sup> The P-values were calculated by linear regression, with the population doubling time as a continuous dependent variable and strain identity as a categorical independent variable, comparing individual strains to the control (R-M-).



**Figure 3-1: RM plasmids and their effect on fitness**

**(A)** The EcoRI and EcoRV RM systems are carried by a pBR322-based plasmid backbone. The  $P_{tet}$  promoter of pBR322 (not shown) was removed to prevent transcription into the RM operon ( $\Delta P_{tet}$ ). Both EcoRI and EcoRV are expressed from their native promoters (Mruk, Liu, Ge, & Kobayashi, 2011; Semenova et al., 2005). The plasmids carry the *bla* gene conferring resistance to ampicillin. The *pMB1 ori* origin of replication is very closely related to *pMB4*, from which EcoRI originates (Betlach et al., 1976). The (R-M+) mutants carry frame-shift mutations, whose locations are denoted by an asterisk above the R gene. The (R-M+) mutants show no restriction activity (Kusano et al., 1995; Nakayama & Kobayashi, 1998). The C protein of EcoRV functions as a transcriptional regulator (Semenova et al., 2005).

**(B)** Competitions in serially transferred cultures between YFP-labeled (R+M+) and CFP-labeled (R-M+) strains carrying EcoRI and EcoRV in rich medium (M9CA). Control represents a competition between YFP and CFP labeled (R-M-) strains to test for the fitness effect of fluorescent labeling. Six independent replicates were performed for each competition. The left plot shows change in the relative ratio between the two strains as  $\ln[(y_t/c_t) * (c_0/y_0)]$ , where  $y_t$  and  $c_t$  represent densities of yellow and cyan cells, respectively at time  $t$  and  $y_0$  and  $c_0$  represent densities at the beginning of the experiment (the strains were initially mixed in approximately 1:1 ratio). For each competition experiment, the selection coefficient was calculated by linear regression of  $\ln[(y_t/c_t) * (c_0/y_0)]$ , against time (Dykhuizen & Hartl, 1983). Average selection coefficients along with the standard deviations are plotted on the right. The dashed line corresponds to no fitness effect ( $s = 0$ ). The results were tested for significance by linear regression, with the selection coefficient as a continuous dependent variable and the type of competition as a categorical independent variable, comparing the result to the control (R-M-) ( $n = 6$  for each competition experiment). No significant fitness effect was detected for EcoRI, nor for EcoRV.

**(C)** Competition experiments in minimal medium (M9, no Casamino acids). The data is plotted as in **Figure 3-1B**. The results were tested for significance by linear regression, with the selection coefficient as a continuous dependent variable and the type of competition as a categorical independent variable, comparing the result to the control (R-M-) ( $F = 34.63$  on 2 and 15 DF,  $p < 10^{-5}$ ). Significant fitness effect of YFP labeling was detected in the control

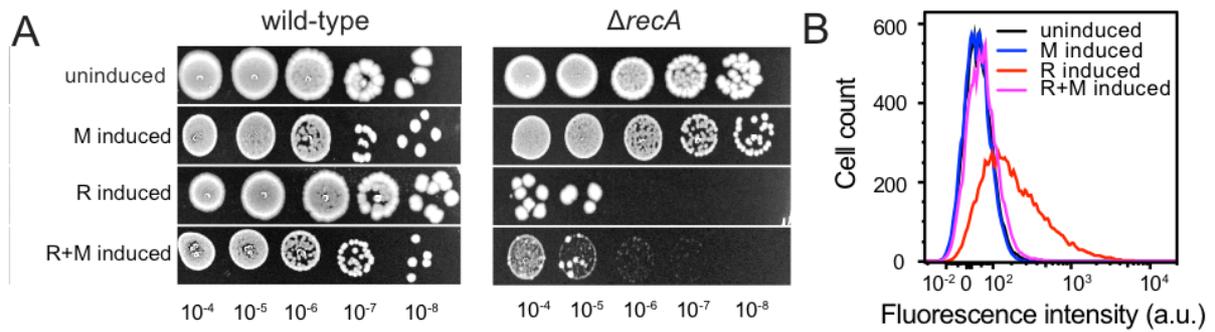
competition experiment ( $s = -0.08 \text{ day}^{-1}$ ,  $p < 10^{-4}$ ). The effect of EcoRV was not significant ( $p = 0.118$ ). Increased cost of EcoRI (R+M+) as compared to the control was observed ( $s = -0.13 \text{ day}^{-1}$ ,  $p < 10^{-5}$ ). At the end of each competition experiment, we tested two colonies of each type for the corresponding restriction phenotype and detected no mutations.

### 3.2.2 EcoRI, but not EcoRV, induces SOS response in a subpopulation of host bacteria

We next investigated whether the increased amount of DNA damage due to EcoRI can be explained by higher frequency of self-restriction as compared to EcoRV. We quantified the fraction of cells suffering from DNA damage in populations carrying the two RM systems using flow cytometry and a reporter strain with a fast-maturing yellow fluorescent protein (YFP) (Nagai et al., 2002) fused to the promoter of *sulA* ( $P_{sulA}$ -*yfp*). SulA is strongly upregulated as a part of the SOS response, a global stress response to DNA damage in *E. coli* (Friedberg, Walker, Siede, & Wood, 2005). Similar  $P_{sulA}$  – based reporters have been previously used to quantify the extent of DNA damage in bacteria (Darmon et al., 2014; Handa, Ichige, Kusano, & Kobayashi, 2000; Pennington & Rosenberg, 2007). Self-restriction provoked by artificially induced imbalance between R and M strongly increased fluorescence of individual cells as a result of SOS response induction (**Figure 3-2B**). When EcoRI (R+M+) was expressed from its native promoters, the population contained more highly fluorescent cells as compared to the controls (**Figure 3-3A**), showing that self-restriction occurred in a subpopulation of cells and induced the SOS response in this subpopulation. No such effect was observed for cells carrying EcoRV.

We quantified the fraction of cells with induced SOS response (SOS-ON) for each strain by first quantifying the fraction of cells with fluorescence above a threshold. The threshold was chosen based on the location at which the cumulative tail probability distribution of the wild-type population changes slope, corresponding to the point at which SOS-OFF and SOS-ON subpopulations begin to overlap (**Figure 3-3B**). Since RecA is necessary for induction of SOS response (Friedberg et al., 2005),  $\Delta recA$  populations did not show such a change in slope. The threshold value was consistent across all samples (**Figure 3-3C**). The wild-type and  $\Delta recA$  control (R-M-) populations contained 0.92% and 0.35% cells with fluorescence above the threshold, respectively (**Figure 3-3D, Table 3-2**). Subtracting the background fraction of cells above the threshold in the  $\Delta recA$  control from the fraction of cells above the

threshold in the wild-type gave an estimate of 0.57% cells being genuinely SOS-ON as a result of spontaneous DNA damage. This is in rough agreement with the previously estimated fraction of 0.9% cells being SOS-ON due to spontaneous DNA damage under slightly different growth conditions (Pennington & Rosenberg, 2007). Using the same method and threshold value, the EcoRI (R+M+) populations contained 0.91% genuine SOS-ON cells, which corresponds to a significant 1.6-fold increase as compared to the EcoRI (R-M+) population (**Figure 3-3D, Table 3-2**). The effect of the EcoRV (R+M+) plasmid on the number of SOS-ON cells was not significant, which is consistent with our observation of EcoRI, but not EcoRV, inducing DNA damage in host bacteria. It is possible that the estimated fraction of cells suffering from self-restriction by EcoRI is an underestimate, since EcoRI, unlike EcoRV, generates cohesive ends that can be directly ligated by DNA ligase before induction of SOS response takes place (Heitman, Zinder, & Model, 1989).



**Figure 3-2: Effects of Imbalanced R and M expression**

**(A)** Overexpression of R.EcoRI does not affect viability in the wild-type, but reduces it in the  $\Delta recA$  background. The effect is rescued by co-expression of M.EcoRI. The rescued colonies are small in size, suggesting that self-restriction still takes place when both enzymes are overexpressed.

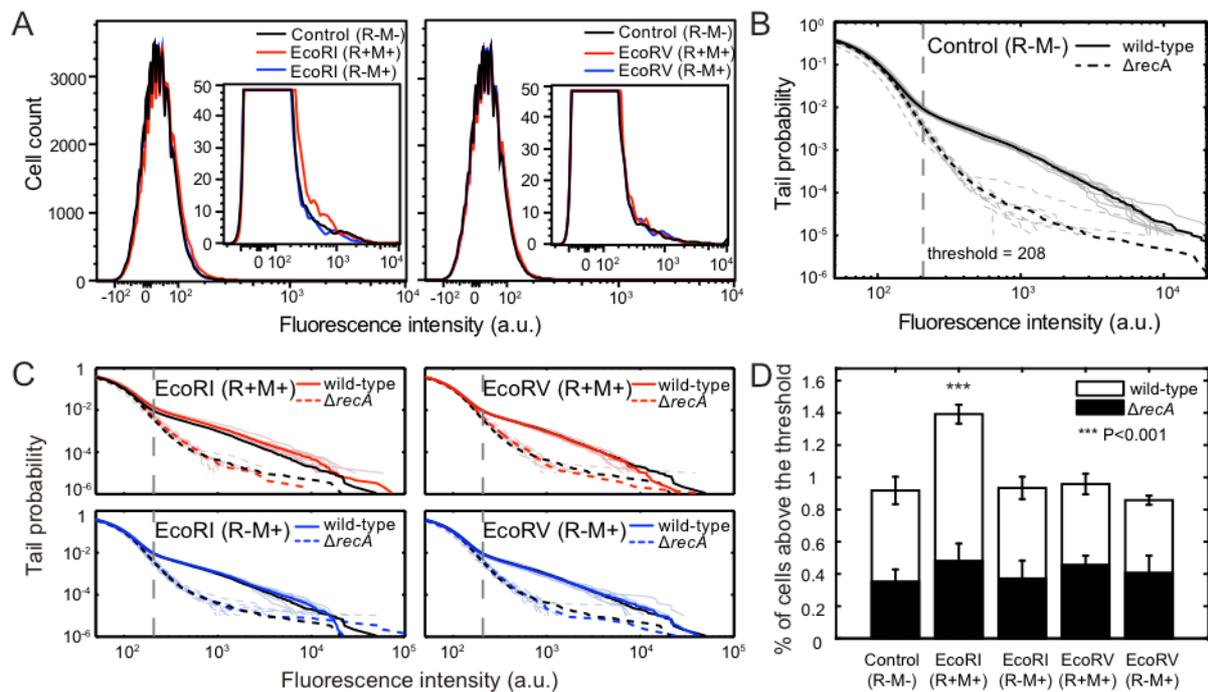
**(B)** Flow cytometry measurements of YFP expressed from the  $P_{sulA}$  promoter in response to R and M imbalance. Overexpression of R.EcoRI in the wild-type increased fluorescence of all cells. Co-induction of R and M largely reversed this effect, suggesting that the increase in SOS response was caused by R and M imbalance and not by toxicity of the inducer or cleavage at non-cognate restriction sites. A small shift in fluorescence was still observable. Both (A) and (B) are representative results of three independent experiments yielding similar results. R.EcoRI and M.EcoRI were expressed from  $P_{tetO-1}$  and  $P_{LacO-1}$  respectively (Lutz & Bujard, 1997), using 10 ng/ml aTc and 100  $\mu$ M IPTG for induction, respectively. The strains contain both pZS\*11-R.EcoRI and pZA32-M.EcoRI inducible plasmids.

**Table 3-2: Fractions of cells above the threshold in populations carrying RM systems**

<b>Host</b>	<b>Plasmid</b>	<b>% of cells above the threshold<sup>a</sup></b>	<b>Standard deviation<sup>a</sup></b>	<b>P-value<sup>a, b</sup></b>
wild-type (MG1655)	Control (R-M-)	0.92	0.08	-
	EcoRI (R+M+)	1.39	0.06	<10 <sup>-5</sup>
	EcoRI (R-M+)	0.93	0.07	0.744
	EcoRV (R+M+)	0.96	0.06	0.434
	EcoRV (R-M+)	0.86	0.03	0.279
$\Delta$ recA	Control (R-M-)	0.35	0.07	-
	EcoRI (R+M+)	0.48	0.11	0.126
	EcoRI (R-M-)	0.37	0.11	0.814
	EcoRV (R+M+)	0.46	0.06	0.206
	EcoRV (R-M+)	0.41	0.11	0.494

<sup>a</sup>. Calculated from three independent experiments, each with three biological replicates.

<sup>b</sup>. The P-values were calculated by linear regression, with the number of cells above the threshold as a continuous dependent variable and strain identity as a categorical independent variable, comparing individual strains to the control (R-M-).



**Figure 3-3: EcoRI, but not EcoRV, induces SOS response in a subpopulation of cells**

**(A)** Representative flow-cytometry histograms of cell fluorescence, corresponding to the YFP expressed from the  $P_{sulA}$  promoter in the presence of EcoRI and EcoRV RM systems expressed from their native promoters. EcoRI increases the number of highly fluorescent cells as a consequence of self-restriction. EcoRV has no effect. The figure inlays show the same data on a rescaled y-axis to emphasize the tail behavior. 100,000 cells were measured in each sample. The x-axes were biexponentially transformed.

**(B)** The threshold value used to quantify the fractions of SOS-ON cells was picked based on the log-log plot of cumulative tail probability vs. fluorescence. 100,000 cells were measured in each sample. Nine biological replicates from three independent experiments are shown as gray lines. Thick black lines represent pooled data. The vertical gray dashed line represents the threshold value = 208 (a.u.).

**(C)** The threshold value = 208 (a.u.) (vertical gray dashed line) was consistent for all strains. The data is plotted as in (B) with pale lines representing nine biological replicates obtained in three independent experiments (100,000 cells measured in each sample). Thick colored lines represent pooled data. The thick black lines represent pooled control (R-M-) populations.

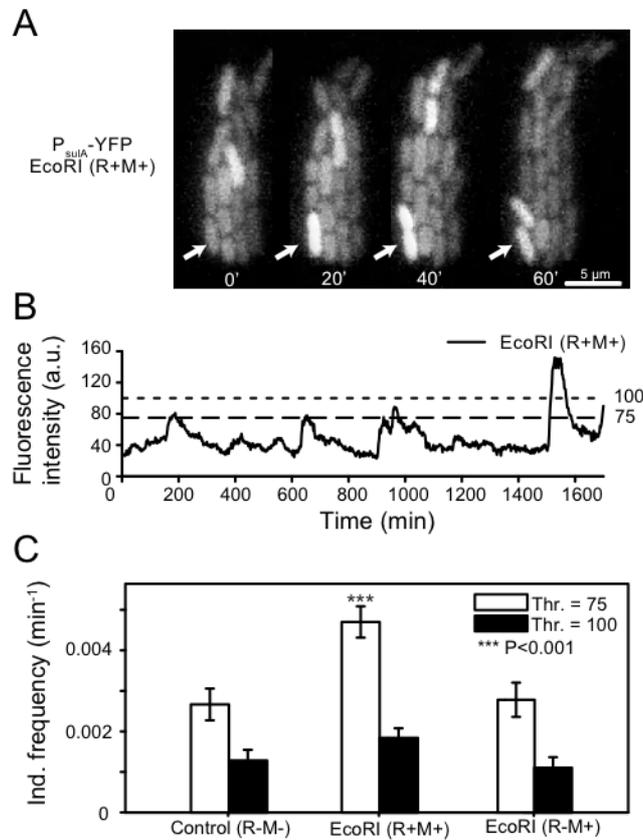
**(D)** Fractions of cells with fluorescence above the threshold in populations carrying RM systems expressed from their native promoters. The error bars represent the standard deviation between three experiment averages (three biological replicates each). Asterisks represent the level of significance. The P-values were calculated by linear regression, with the number of cells above the threshold as a continuous dependent variable and strain identity as a categorical independent variable, comparing individual strains to the control (R-M-). The number of genuine SOS-ON cells can be obtained by subtracting the background fraction of cells above the threshold in the  $\Delta recA$  control from the fraction of cells above the threshold in the wild-type.

### 3.2.3 SOS response is stochastically and dynamically turned on and off in cells suffering from self-restriction

To monitor the fate of cells undergoing self-restriction, we observed single cells carrying EcoRI in real time using fluorescence long-term time-lapse microscopy. We measured the levels of  $P_{sulA-yfp}$  expression in single cells growing in steady state inside a microfluidic device (**Figure 3-4A**). Since RecA is necessary for SOS induction (Friedberg et al., 2005), we first determined the threshold of fluorescence above which cells are evaluated as SOS-ON using a  $\Delta recA$  control. Cellular fluorescence of  $\Delta recA$  cells fluctuated due to noise, but no sharp increase was observed in contrast to wild-type cells (**Figure 3-5A**). We set the threshold accordingly and calculated the frequency at which fluorescence intensity of wild-type cells carrying the EcoRI RM system crossed the threshold (**Figure 3-4B, Table 3-2**). At the threshold value of 75 (a.u.), the wild-type restriction-deficient strains (control (R-M-) and EcoRI (R-M+)) displayed nearly identical frequency of SOS induction:  $(2.7 \pm 0.4) \times 10^{-3} \text{ min}^{-1}$  and  $(2.8 \pm 0.4) \times 10^{-3} \text{ min}^{-1}$ , respectively, as a result of spontaneous DNA damage. The EcoRI (R+M+) strain induced SOS response at the rate of  $(4.7 \pm 0.4) \times 10^{-3} \text{ min}^{-1}$ , which corresponds to a 1.7-fold increase (**Figure 3-4C**). In total, we observed 0.53% and 0.94% cells being SOS-ON in the control (R-M-) and EcoRI (R+M+) populations respectively (**Table 3-3**), which is in agreement with the flow-cytometry experiments. Using a threshold value of 100 (a.u.) did not affect the result qualitatively (1.4-fold increase in SOS induction frequency of EcoRI (R+M+) cells).

Interestingly, the SOS-ON cells in the EcoRI (R+M+) as well as EcoRI (R-M+) and control (R-M-) populations returned rapidly to the SOS-OFF state and continued to grow and divide normally (**Figure 3-4A**).  $P_{sulA-yfp}$  induction was not associated with filamentation or cell death and the level of SOS response induction did not correlate with the single-cell elongation rate (**Figure 3-5B**) or generation time (**Figure 3-5C**). These results indicated that under our experimental conditions, wild-type cells growing in steady state repair the DNA damage caused by both self-restriction and spontaneous DNA damage with high efficiency and thus remain in the growing population. In our experiments, self-restriction occurs during stable maintenance of the RM system and induces the SOS response only transiently, without affecting viability of individual cells. This stands in sharp contrast to the previously

described process of post-segregational killing (Naito et al., 1995), which occurs when intracellular levels of R and M are irreversibly disturbed by dilution following gene loss (Ichige & Kobayashi, 2005) and ultimately leads to cell death (Asakura & Kobayashi, 2009).

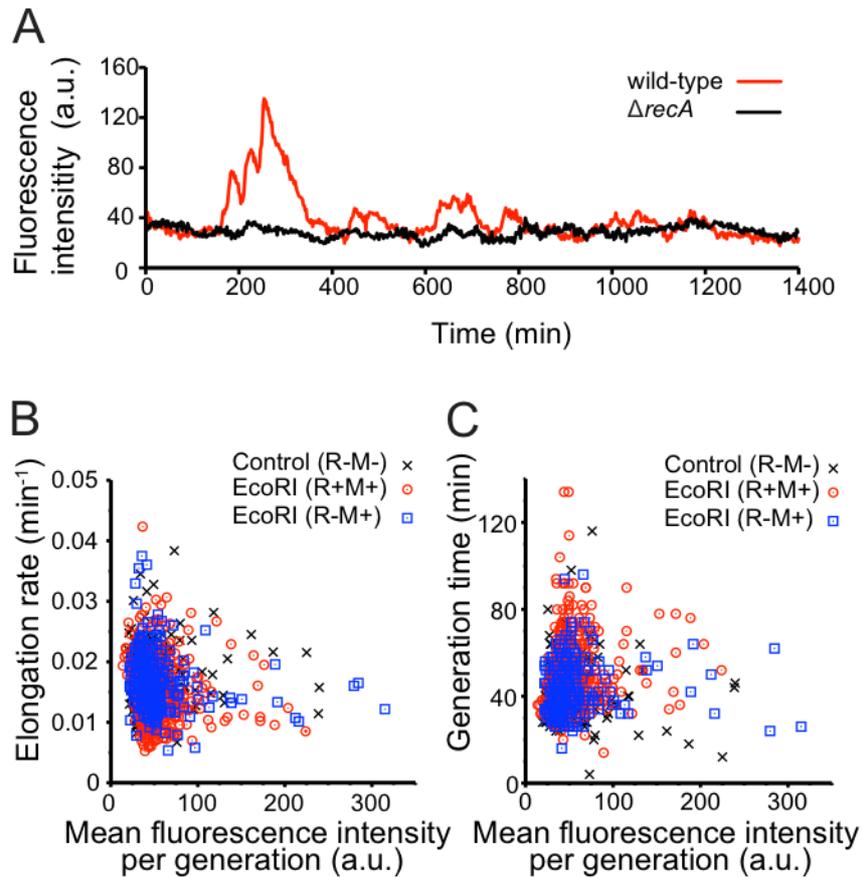


**Figure 3-4: Real-time dynamics of self-restriction by EcoRI in single cells**

**(A)** Representative time-lapse images showing spontaneous temporary induction of YFP expressed from the  $P_{sulA}$  promoter in wild-type cells carrying the EcoRI (R+M+) RM system. Cells growing inside a microfluidic device show spontaneous induction of SOS response, followed by dilution of YFP due to cell division. White arrows point to the cell that underwent SOS induction and subsequent cell division.

**(B)** Representative single cell lineage showing dynamics of YFP expression from the  $P_{sulA}$  promoter of the EcoRI (R+M+) strain. The fluorescence intensity at each time point is evaluated as the mean pixel brightness within a region corresponding to a single cell subtracted by the background brightness. The horizontal dashed lines indicate the threshold levels used to calculate SOS induction frequencies.

**(C)** SOS induction frequencies calculated as the total counts of fluorescence intensity crossing the respective threshold values divided by the total time-length of all the branches in the lineage trees, on which we found 463 control (R-M-), 866 EcoRI (R+M+), and 465 EcoRI (R-M+) cells including those flown away from the growth channels before division. The error bars indicate the binomial errors in calculating the induction frequency.



**Figure 3-5: Frequency of SOS-induction and its effects growth of individual bacteria**

**(A)** Representative single cell lineages showing  $P_{sulA}$ - $yfp$  expression dynamics of the wild-type strain and the  $\Delta recA$  strain (both (R-M-)) for  $\sim 35$  generations. The fluorescence intensity at each time point is evaluated as the mean pixel brightness within a region corresponding to a single cell subtracted by the background brightness. Fluorescence of the  $\Delta recA$  cells fluctuated in the range of 15-45 (a.u.) due to noise, but no sharp increase in fluorescence was observed, as in the case for the wild-type cells.

**(B)** Relationship between the elongation rate of single cells and the mean  $P_{sulA}$ - $yfp$  fluorescence intensity per generation. Elongation rate is evaluate as:  $\log(\text{Final cell size}/\text{Initial cell size})/\text{Generation time}$ . The Pearson correlation coefficients ( $r$ ) for the three strains are: Control (R-M-):  $0.03 \pm 0.06$  ( $n = 225$ ), EcoRI (R+M+):  $-0.08 \pm 0.05$  ( $n = 423$ ), EcoRI (R-M+):  $-0.20 \pm 0.06$  ( $n = 236$ ).

**(C)** Relationship between the generation time of single cells and mean  $P_{sulA}$ - $yfp$  fluorescence intensity per generation. The Pearson correlation coefficients ( $r$ ) for the three strains are: Control (R-M-):  $-0.04 \pm 0.07$  ( $n = 225$ ), EcoRI (R+M+):  $0.12 \pm 0.05$  ( $n = 423$ ), EcoRI (R-M+):  $0.06 \pm 0.07$  ( $n = 236$ ).

**Table 3-3: Induction frequencies of SOS response inside microfluidic device.  
Threshold = 75 (a.u.)**

Plasmid	Total time points (a)	Induction time points (b)	Induction frequency (b)/(a) (%)	Induction frequency (b)/(a)/ $\Delta t$ ( $10^{-3} \text{ min}^{-1}$ ) <sup>a</sup>	P-value <sup>b</sup>
Control (R-M-)	8,629	46	0.533	2.67	-
EcoRI (R+M+)	15,533	146	0.940	4.70	<10 <sup>-3</sup>
EcoRI (R-M-)	7,735	43	0.556	2.78	0.927

**Threshold = 100 (a.u.)**

Pladmid	Total time points (a)	Induction time points (b)	Induction frequency (b)/(a) (%)	Induction frequency (b)/(a)/ $\Delta t$ ( $10^{-3} \text{ min}^{-1}$ ) <sup>*</sup>	P-value <sup>†</sup>
Control (R-M-)	8,629	22	0.255	1.27	-
EcoRI (R+M+)	15,533	57	0.367	1.83	0.179
EcoRI (R-M-)	7,735	17	0.220	1.10	0.764

<sup>a</sup>  $\Delta t$  is the time-lapse interval, which is 2 min in all the microfluidic measurements.

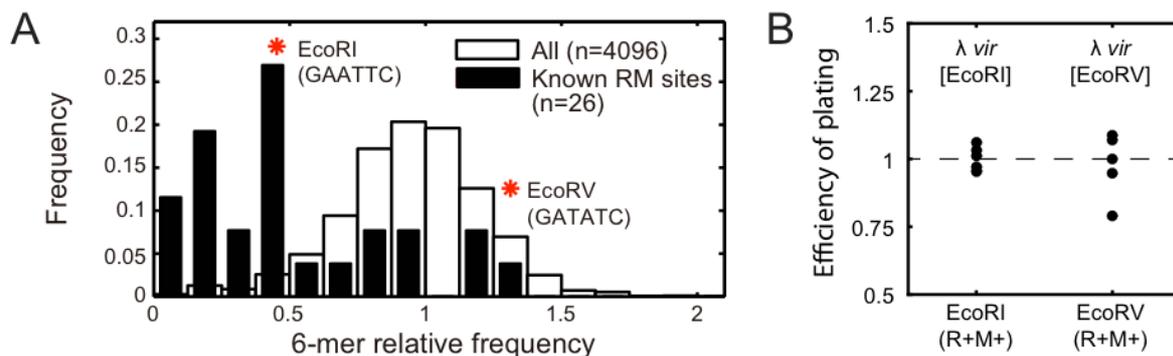
<sup>b</sup> P-values are from the chi-square test with the control (R-M-) based on the contingency table with the entries of (b) and (a)-(b).

### 3.2.4 The rate of self-restriction is higher for more efficient RM systems and can be increased by stochastic imbalance between R and M

Our experiments show that the probability of self-restriction is higher for EcoRI than EcoRV. Interestingly, this does not correspond to the number of restriction sites that are potential targets for self-restriction in the genome of *E. coli* MG1655 (599 GAATC for EcoRI and 1,888 GATATC for EcoRV), but agrees with the estimated levels of restriction site avoidance (**Figure 3-6A**). The EcoRI restriction site frequency is reduced by 50% from its expected value, while the EcoRV site is slightly enriched compared to expectation. This difference in EcoRI and EcoRV restriction site frequencies was previously noticed (Gelfand & Koonin, 1997), although not explained. We hypothesized that the difference in self-restriction rate between the two RM systems results from an intrinsic difference in restriction efficiency per single restriction site (probability that a restriction site is cleaved before methylation). We tested this hypothesis by measuring the efficiency of EcoRI and EcoRV in preventing infection by unmethylated bacteriophage  $\lambda$  *vir*. The efficiency of plating (*eop*), reflecting the probability of  $\lambda$  *vir* escaping restriction, was  $1.6 \times 10^{-5}$  and  $2.7 \times 10^{-8}$  for EcoRI and EcoRV, respectively (**Figure 3-7A**). Assuming that a phage with  $n$  restriction sites (5 for EcoRI and 22 for EcoRV in  $\lambda$  *vir* (R. Roberts, Vincze, Posfai, & Macelis, 2015)) escapes restriction when all its restriction sites are methylated before cleavage occurs (Enikeeva, Severinov, & Gelfand, 2010), the restriction efficiency per restriction site is given by  $1 - \sqrt[n]{eop}$ . In agreement with our hypothesis, the restriction efficiency is significantly higher for EcoRI than EcoRV (**Figure 3-7B**). Neither EcoRI, nor EcoRV restricted fully methylated  $\lambda$  *vir* (**Figure 3-6B**), indicating that under our experimental conditions, EcoRI and EcoRV do not cleave DNA at methylated or non-cognate restriction sites (Vasu, Nagamalleswari, & Nagaraja, 2012). Their different restriction efficiencies thus likely reflect a difference in R and M gene expression levels and enzymatic activities.

These results suggest that RM systems with higher per site probability of cleavage are more likely to cause self-restriction and lead to stronger restriction site avoidance. Although a variety of gene regulatory mechanisms are known to maintain well-balanced levels of R and M expression (Mruk & Blumenthal, 2008; Mruk & Kobayashi, 2014; Mruk et al., 2011; Nagornykh et al., 2008; Semenova et al., 2005), stochastic events occurring at the level of

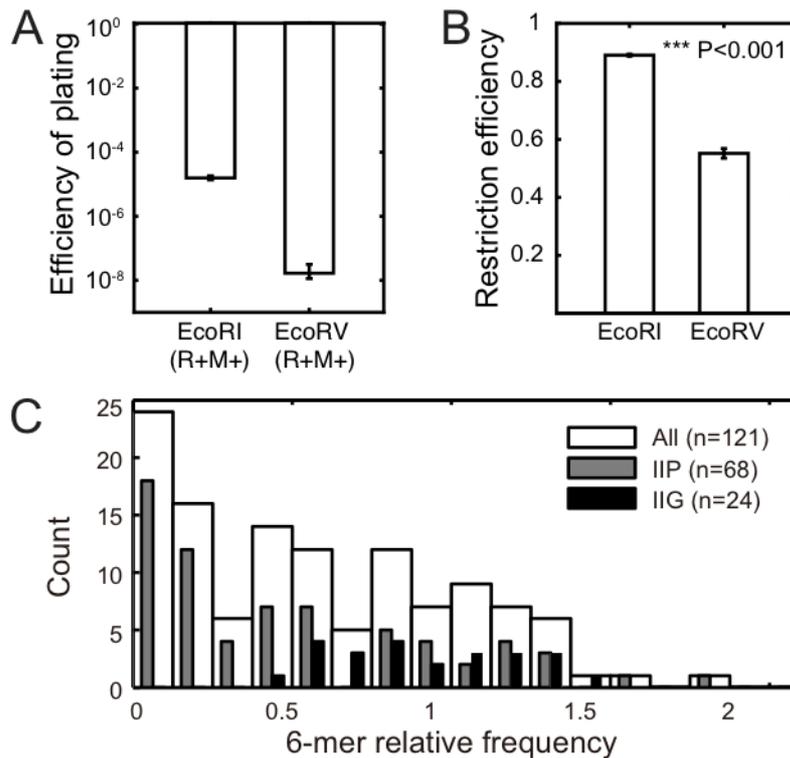
the single cell, such as stochastic gene expression (Elowitz et al., 2002) or protein partitioning at cell division (Huh & Paulsson, 2011), might occasionally disrupt this balance and contribute to the overall rate of self-restriction. In support of this hypothesis, we found that restriction sites of type IIP RM systems, in which R and M are structurally and functionally independent enzymes, were on average more avoided and exhibited a wider range of genomic frequencies than restriction sites of type IIG RM systems, in which R and M are fused into a single bifunctional polypeptide (R. Roberts, 2003) (**Figure 3-7C**). The direct linkage of R and M will minimize the probability of stochastic R and M imbalance due to fluctuations of individual components, which is expected to reduce self-restriction rates. In contrast, self-restriction in type IIP systems (which include EcoRI and EcoRV) can result from differences in expression levels and enzymatic activities of R and M as well as stochastic imbalance between their concentrations. This additional source of variance in type IIP systems is consistent with the significantly wider range of their restriction site frequencies and higher avoidance on average.



**Figure 3-6: Restriction site avoidance and enzyme specificity**

**(A)** Avoidance of restriction sites in the genome of *Escherichia coli*. Histogram of relative frequencies of all 6 bp words in the coding regions of *E. coli* MG1665 (NC\_000913) is shown as white bars. The actual count for each word is normalized by its expected value computed from partially randomized coding sequences. The histogram of known 6 bp restriction sites of type II R enzymes in *E. coli* is shown in black bars. Asterisks indicate relative frequencies of the restriction sites for EcoRI and EcoRV. The EcoRI restriction site frequency is reduced by 50% from its expected value, while the EcoRV site is slightly enriched compared to expectation.

**(B)** EcoRI and EcoRV do not cleave DNA at methylated or non-cognate restriction sites. Fully methylated EcoRI and EcoRV  $\lambda$  vir lysates were prepared by growing  $\lambda$  vir on lawns of MP062 (pZA32-M.EcoRI) and MP062 (pZA32-M.EcoRV). Expression of the methyltransferases was induced by adding 100  $\mu$ M IPTG into the soft agar. Six independent measurements are displayed as individual data points. The titer of the lysates was obtained by plating on control (R-M-) lawns in six independent experiments. The dashed line corresponds to the efficiency of plating = 1, corresponding to no restriction. Neither EcoRI, nor EcoRV restricted fully methylated  $\lambda$  vir, suggesting that under our experimental conditions, EcoRI and EcoRV do not cleave methylated or non-cognate restriction sites.



**Figure 3-7: Identifying determinants of self-restriction**

**(A)** Efficiency of plating of  $\lambda$  *vir* on the lawns of cells with respective RM systems, corresponding to the probability that the phage will not be restricted by the RM system.  $\lambda$  *vir* contains 5 EcoRI and 22 EcoRV restriction sites. The error bars represent standard deviation calculated from four independent experiments.

**(B)** Efficiencies of restriction per restriction site. The values were calculated from the data shown in **Figure 3-7A**. The error bars represent standard deviation calculated from four independent experiments. P-values were calculated with the t-test (T=39.6 on 6 DF, P<10<sup>-4</sup>).

**(C)** For all predicted 6-cutter RM systems in the stringent set, the distribution of relative frequencies of putative restriction sites in the assigned host genomes is shown as white bars. Relative frequencies of 6 bp restriction sites of type IIG (where R and M are fused into a single bifunctional unit) and the canonical type IIP (where R and M are structurally and functionally independent) RM systems are shown as black and gray bars respectively. The 29 remaining RM systems are of type IIS, type IIF or solitary endonucleases and are not shown in separate bars.

### **3.3 Discussion**

Our finding that a more efficient RM system exhibits a higher self-restriction rate is indicative of an evolutionary tradeoff between enhanced protection against exogenous DNA and increased autoimmunity. Evolution of restriction site avoidance in a genome mitigates the long-term cost of an RM system, which was previously estimated for 6-cutter enzymes to be  $10^{-5}$  -  $10^{-4}$  per generation at mutation-selection balance (Qian & Kussell, 2012). While we did not observe a measurable fitness cost of self-restriction under standard conditions, we did observe a noticeable fitness cost when resources were limited. The long-term cost of RM systems in natural populations will thus depend both on molecular properties of individual RM systems and on environmental determinants. These results are in accord with studies showing that other phage resistance mechanisms such as clustered regularly interspaced short palindromic repeats CRISPR-associated (CRISPR-cas) systems (Vale et al., 2015; Westra et al., 2015), abortive infection (Berngruber, Lion, & Gandon, 2013; Refardt, Bergmiller, & Kümmerli, 2013) or envelope resistance (Lenski, 1988; Lenski & Levin, 1985) come with a cost for the host cell, and that the cost of immunity can be accentuated in environments with limited resource availability in bacteria (Gómez & Buckling, 2011), as well as in higher organisms (Boots, 2011).

The ability to discriminate self from non-self is a crucial property of all immune systems (Marraffini & Sontheimer, 2010), and failure to do so leads to pathogen tolerance or autoimmunity (Goldberg & Marraffini, 2015). In this work, we show that similarly to more complex immune systems, autoimmunity due to RM systems affects a small number of individuals in a population. RM systems are extremely abundant in prokaryotes (Oliveira et al., 2014; Vasu & Nagaraja, 2013) and likely play an important role in their ecology and evolution. Understanding the costs and benefits associated with RM systems is crucial to fully evaluate this role (Korona & Levin, 1993). RM systems can protect their hosts from parasites (Arber & Dussoix, 1962), but also act parasitically (Naito et al., 1995). While they act as a barrier to horizontal gene transfer (Corvaglia et al., 2010; Murray, 2002), they themselves are often mobile (Furuta & Kobayashi, 2012) and can even promote DNA recombination (Arber, 2000). Here, we describe a new type of interaction between RM systems and their hosts – a primitive form of bacterial autoimmunity. As a downside of an

immunity mechanism based on a balance between individual components, bacterial self-restriction exemplifies a link between stochastic events occurring at the level of single individuals and the evolution of bacterial genomes.

### 3.4 **Material and Methods**

#### 3.4.1 **Bacterial strains, plasmids and growth conditions**

The EcoRI and EcoRV RM systems are encoded on the pBR322 backbone (AmpR, 15-20 copies per cell) and are expressed in their native configuration from their native promoters (**Figure 3-1A**). The origin of replication of pBR322 (pMB1) is closely related to pMB4 from which EcoRI originates (Betlach et al., 1976). To prevent post-segregational killing (Handa & Kobayashi, 1999), we constantly selected for plasmid maintenance, although pBR322-derived plasmids have been shown to be stable over hundreds of generations in the absence of selection pressure (Chiang & Bremer, 1988). The  $P_{sulA}$ -*yfp* reporter was integrated in the HK022 attachment site in the host chromosome. M9CA medium (1x M9 salts, 0.4% glucose, 2mM MgSO<sub>4</sub>, 0.1mM CaCl<sub>2</sub>, 0.2% Casamino acids) was used for bacterial growth unless otherwise stated. The medium was supplemented with antibiotics when needed at following concentrations: Ampicillin 100 µg/ml, Chloramphenicol 15 µg/ml, Kanamycin 50 µg/ml. Liquid bacterial cultures were propagated in the dark, at 37°C with vigorous shaking.

#### 3.4.2 **Population doubling time measurements and spotting assay**

For estimation of population doubling times, overnight cultures started from individual colonies were diluted 1:250 in a flat-bottom 96-well plate into fresh medium (200 µl total volume). The plate was continuously shaken inside the Synergy H1 Multi-Mode Reader and OD600 was measured at ten-minute intervals for 10 hours. Growth-rates were calculated from the background-subtracted values of OD600 as the slope of ln(OD600) vs. time during 90 minutes of exponential growth. Population doubling times were calculated as ln(2)/growth rate. Outmost wells were used for background subtraction. We repeated the growth-rate experiment in 5 ml batch cultures and obtained consistent results (not shown) with these presented in **Table 3-1**. For the spotting assay, overnight cultures were serially diluted in SM buffer and spotted (10 µl) on M9 plates (without Casamino acids) supplemented with respective antibiotics and inducers. Plates were incubated at 37°C for 48 h.

### 3.4.3 Flow cytometry measurements and analysis

Overnight cultures started from single colonies were diluted 1:1,000 into 2 ml fresh medium and grown for 4 hours to reach the exponential phase. The samples were diluted 1:100 into filtered (0.22  $\mu\text{m}$ ) medium immediately prior to measurement. Fluorescence was measured using the BD FACSCanto™ II system in the FITC channel. Fluorescence intensity data was exported into the ASCII format using FlowJo (Tree Star) and analyzed in MATLAB. The fractions of SOS-ON cells were obtained using a single threshold value determined by inspection of the log-log plots of cumulative tail probability vs. fluorescence (**Figure 3-3B**). The cumulative tail probability is defined as the cumulative probability of cells with fluorescence higher than  $x$ . We identified the vicinity of a kink in the wild-type (R-M-) control, and used it as a natural threshold value for quantifying fractions of SOS-ON cells for all the samples analyzed. The fraction of genuine SOS on cells was obtained by subtracting the fraction of cells above the threshold in the  $\Delta\text{recA}$  population, from the fraction of cells above the threshold in the wild-type.

### 3.4.4 Measuring restriction efficiency

Efficiency of plating ( $eop$ ) was determined by plating serially diluted  $\lambda$  *vir* lysate (grown on MG1655) on lawns of bacteria with corresponding plasmids. Overnight cultures were mixed with serial dilutions of the lysate in 3 ml of phage soft agar (1% tryptone, 0.1% yeast extract, 0.8% NaCl, 0.7% Agar, 0.01% glucose, 0.2 mM  $\text{CaCl}_2$ ) and plated on phage plates (1% tryptone, 0.1% yeast extract, 0.8% NaCl, 1% Agar, 0.01% glucose, 0.2 mM  $\text{CaCl}_2$ ). The number of plaque-forming units  $n_{pfu}$  per ml of the lysate was calculated for each strains as:  $n_{\text{plaques}} \times \text{dilution}$ . Efficiency of plating ( $eop$ ), which corresponds to the probability that all phage restriction sites are methylated before cleavage, was calculated as  $eop = n_{pfu\_sample} / n_{pfu\_control}$ , where  $n_{pfu\_control}$  is the number of plaque-forming units obtained on the control (R-M-) lawn. For a phage with  $n$  restriction sites, the probability that a single restriction site is methylated before cleavage is given by  $\sqrt[n]{eop}$ . Restriction efficiency (the probability that a single restriction site is cleaved before methylation) is then given by  $1 - \sqrt[n]{eop}$ . We measured no significant decrease in  $eop$  for the EcoRI (R-M+) ( $0.97 \pm 0.07$ ) and EcoRV (R-M+) ( $0.99 \pm 0.08$ ) strains ( $\text{mean} \pm \text{SD}$ ,  $n=4$ ).

### **3.4.5 Microscopy fluorescence measurements**

A Nikon Ti-E microscope equipped with a thermostat chamber (TIZHB, Tokai Hit), 100× oil immersion objective (Plan Apo λ, N.A. 1.45, Nikon), cooled CCD camera (ORCA-Flash, Hamamatsu Photonics) and LED excitation light source (DC2100, Thorlabs) was used for the microscopy fluorescence measurements. The microscope was controlled by micromanager (<https://micro-manager.org>). The cells were grown in a microfluidic device similar to the ‘mother machine’ (P. Wang et al., 2010). The growth channel dimensions were 5 mm(W)×15 mm(L)×1 mm(D), which allowed to stably harbor approximately 10 cells. Overnight cultures were diluted 1:400 into fresh medium and grown at 37°C for 4 hours to reach exponential phase. The cultures were centrifuged to obtain the cell suspension of OD ~2.0 and injected into the device with a 1-ml syringe. After the 30-minute incubation at 37°C to load the cells into the growth channels, fresh pre-warmed medium was flown at the rate of 5 ml/h for 5 minutes to remove the cells outside the growth channels. The flow rate was fixed to 1 ml/h throughout the measurements. Phase, YFP fluorescence and RFP fluorescence images were taken simultaneously at 2-min time-lapse interval. Multiple growth channels were monitored simultaneously in a single experiment. We used a custom macro of ImageJ (<http://imagej.nih.gov/ij/>) for the image analysis and a custom C-program for the data analysis.

### **3.4.6 Identifying putative R and M genes and assigning properties**

Reference sequences of experimentally validated type II R and M proteins with 4-6 bp restriction sites were downloaded from the ReBase Gold Standard Set (R. Roberts et al., 2015). Using these as templates, we searched for potential R and M genes in all annotated full-length prokaryotic chromosome sequences (downloaded from the Genbank RefSeq prokaryotic collection) by BLASTP with the criteria of e-value < 10<sup>-10</sup>. The list of hits was filtered for alignment length within ±20% of the template length and protein sequence similarity > 50%. These hits constituted our full set of R and M predictions. To construct a stringent set of potential homologues, 59 non-redundant 6-cutter R enzymes that had matches with protein similarity > 80% were considered the core set of “prototypes” with the source genomes assigned as their hosts. For each prototype enzyme, its (forward) top hit in

each genome was used as the template for a (reverse) BLASTP search against all genes in the prototype's host genome. If the prototype ranked #1 among all the reverse hits, the corresponding forward top hit was included in the stringent dataset. A stringent dataset for the M enzyme was constructed in the same way from 73 prototype M enzymes. Target specificity was assigned to all hits according to their prototype enzymes. A RM system (R-M pair) was identified when two R and M predictions sit within +/- 2 genes in the same genome. Subtype was assigned according to prototypes. For the C-controlled RM systems, a matching control gene also had to be present to call the assignment. Since there is essentially no information on the activity of these systems in their respective hosts, we cannot rule out the possibility that a fraction of RM systems in our set are no longer functional or have changed target specificity. For each R enzyme in the stringent dataset, we measured the frequency  $q_w$  of its assigned restriction site  $w$  by a sliding window across all coding regions of the host genome. To account for word frequency deviations due to protein coding constraints, we randomized the original sequence by shuffling the synonymous codons genome-wide, preserving the amino acid sequences and codon usage biases. The expected word frequency  $q_w^0$  is obtained by averaging frequencies over 1,000 randomizations of the entire coding sequence. The relative frequency is defined as  $q_w/q_w^0$ . In cases where the target specificity is degenerate, e.g., ARCGYT (R = G/A, Y = C/T), all words compatible with the pattern are assessed separately.

### **3.4.7 Competition experiments**

The EcoRI (R+M+) and (R-M+) and EcoRV (R+M+) and (R-M+) plasmids were transformed into MG1655 strains expressing Venus YFP and Cerulean CFP from the  $\lambda$  P<sub>R</sub> promoter, integrated in the *attP21* site on the chromosome of the host. A single colony for each host strain/plasmid combination was picked from a freshly streaked plate and grown overnight in 2 ml of corresponding medium supplied with ampicillin. The strains were mixed in approximately 1:1 ratio and diluted 1:1,000 into 2 ml of fresh medium and grown for 24 hours. Cultures were diluted every 24 hours by 1:1,000 into 2 ml of fresh medium. We constantly selected for the plasmid maintenance by supplying the medium with ampicillin. The densities of individual types were measured by sampling 100  $\mu$ l at the end of each transfer, diluting the sample in SM buffer and plating on LB plates such that 100-200

colonies per plate were obtained. Two measurements (independent platings) for each sample were averaged to obtain the density of each type. Fluorescent images of the plates were taken using a custom build microscope (Chait, Shrestha, Shah, Michel, & Kishony, 2010) and colonies of each type were counted manually. The selection coefficients  $s$  were calculated by linear regression of  $\ln[(y_t/c_t) * (c_0/y_0)]$ , against time (Dykhuizen & Hartl, 1983), where  $y_t$  and  $c_t$  represent densities of yellow and cyan cells, respectively at time  $t$  and  $y_0$  and  $c_0$  represent densities at the beginning of the experiment.

### 3.4.8 List of strains and plasmids

Name	Genotype	Source, reference
DH5 $\alpha$	<i>F</i> <sup>-</sup> , $\Phi$ 80 <i>lacZ</i> $\Delta$ M15, $\Delta$ ( <i>lacZYA-argF</i> ), U169, <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> ( <i>rK</i> <sup>-</sup> , <i>mK</i> <sup>+</sup> ), <i>phoA</i> <i>supE44</i> , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i> , $\lambda$ <sup>-</sup>	Lab collection
DH5 $\alpha$ $\lambda$ <i>pir</i> <sup>+</sup>	<i>F</i> <sup>-</sup> , $\Phi$ 80 <i>lacZ</i> $\Delta$ M15, $\Delta$ ( <i>lacZYA-argF</i> ), U169, <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> ( <i>rK</i> <sup>-</sup> , <i>mK</i> <sup>+</sup> ), <i>phoA</i> <i>supE44</i> , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i> , $\lambda$ <i>pir</i> <sup>+</sup>	Lab collection
Frag1B	<i>F</i> <sup>-</sup> , <i>rha</i> <sup>-</sup> , <i>thi</i> , <i>gal</i> , <i>lacZ.AM</i> , , $\Delta$ <i>att</i> $\lambda$ ::( <i>P</i> <sub>N25</sub> - <i>tetR</i> , <i>P</i> <sub><i>lacI</i><sup>q</sup></sub> - <i>lacI</i> , <i>Sp</i> <sup>R</sup> )	Lab collection
MG1655	<i>F</i> <sup>-</sup> , $\lambda$ <sup>-</sup> , <i>ilvG</i> <sup>-</sup> , <i>rfb-50</i> , <i>rph-1</i>	Lab collection
MP060	MG1655, $\Delta$ <i>attHK022</i> ::( <i>P</i> <sub><i>sulA</i></sub> - <i>yfp</i> )	This work
MP062	MG1655, $\Delta$ <i>attHK022</i> ::( <i>P</i> <sub><i>sulA</i></sub> - <i>yfp</i> ), $\Delta$ <i>att</i> $\lambda$ ::( <i>P</i> <sub>N25</sub> - <i>tetR</i> , <i>P</i> <sub><i>lacI</i><sup>q</sup></sub> - <i>lacI</i> , <i>Sp</i> <sup>R</sup> )	This work
MP064	MG1655, $\Delta$ <i>attHK022</i> ::( <i>P</i> <sub><i>sulA</i></sub> - <i>yfp</i> ), $\Delta$ <i>recA</i>	This work
MP066	MG1655, $\Delta$ <i>attHK022</i> ::( <i>P</i> <sub><i>sulA</i></sub> - <i>yfp</i> ), $\Delta$ <i>attP21</i> ::( <i>P</i> <sub>R</sub> - <i>mCherry</i> )	This work
MP068	MG1655, $\Delta$ <i>attHK022</i> ::( <i>P</i> <sub><i>sulA</i></sub> - <i>yfp</i> ), $\Delta$ <i>attP21</i> ::( <i>P</i> <sub>R</sub> - <i>mCherry</i> ), $\Delta$ <i>recA</i>	This work
MP070	MG1655, $\Delta$ <i>recA</i>	This work
MP074	MG1655, $\Delta$ <i>attHK022</i> ::( <i>P</i> <sub><i>sulA</i></sub> - <i>yfp</i> ), $\Delta$ <i>att</i> $\lambda$ ::( <i>P</i> <sub>N25</sub> - <i>tetR</i> , <i>P</i> <sub><i>lacI</i><sup>q</sup></sub> - <i>lacI</i> , <i>Sp</i> <sup>R</sup> ), $\Delta$ <i>recA</i>	This work
MG1655 <i>P</i> <sub>R</sub> -Venus	MG1655, $\Delta$ <i>attP21</i> ::( <i>P</i> <sub>R</sub> -Venus)	Lab collection
MG1655 <i>P</i> <sub>R</sub> -Cerulean	MG1655, $\Delta$ <i>attP21</i> ::( <i>P</i> <sub>R</sub> -Cerulean)	Lab collection
pBR322	<i>pMB1 ori</i> , <i>Amp</i> <sup>R</sup> , <i>Tet</i> <sup>R</sup>	NEB
pIK166	<i>pMB1 ori</i> , <i>Amp</i> <sup>R</sup> , <i>EcoRI</i> (R+M+)	Ichizo Kobayashi
pIK167	<i>pMB1 ori</i> , <i>Amp</i> <sup>R</sup> , <i>EcoRI</i> (R-M+)	Ichizo Kobayashi
pYNEC107	<i>pMB1 ori</i> , <i>Amp</i> <sup>R</sup> , <i>EcoRV</i> (R+M+)	Ichizo Kobayashi
pYNEC117	<i>pMB1 ori</i> , <i>Amp</i> <sup>R</sup> , <i>EcoRV</i> (R-M+)	Ichizo Kobayashi
pBR322 $\Delta$ <i>P</i> <sub>tet</sub>	<i>pMB1 ori</i> , <i>Amp</i> <sup>R</sup> , $\Delta$ <i>P</i> <sub>tet</sub>	This work
pAH68- <i>frt</i> - <i>cam</i>	<i>R6K ori</i> , <i>Cam</i> <sup>R</sup> - <i>frt</i> , <i>attPHK022</i> , <i>yfp</i>	Lab collection
pMP017	<i>R6K ori</i> , <i>Cam</i> <sup>R</sup> - <i>frt</i> , <i>attPHK022</i> , <i>P</i> <sub><i>sulA</i></sub> - <i>yfp</i>	This work
pBR322 $\Delta$ <i>P</i> <sub>tet</sub> <i>EcoRI</i> (R+M+)	<i>pMB1 ori</i> , <i>Amp</i> <sup>R</sup> , $\Delta$ <i>P</i> <sub>tet</sub> , <i>EcoRI</i> (R+M+)	This work
pBR322 $\Delta$ <i>P</i> <sub>tet</sub> <i>EcoRI</i> (R-M+)	<i>pMB1 ori</i> , <i>Amp</i> <sup>R</sup> , $\Delta$ <i>P</i> <sub>tet</sub> , <i>EcoRI</i> (R-M+)	This work
pBR322 $\Delta$ <i>P</i> <sub>tet</sub> <i>EcoRV</i> (R+M+)	<i>pMB1 ori</i> , <i>Amp</i> <sup>R</sup> , $\Delta$ <i>P</i> <sub>tet</sub> , <i>EcoRV</i> (R+M+)	This work
pBR322 $\Delta$ <i>P</i> <sub>tet</sub> <i>EcoRV</i> (R-M+)	<i>pMB1 ori</i> , <i>Amp</i> <sup>R</sup> , $\Delta$ <i>P</i> <sub>tet</sub> , <i>EcoRV</i> (R-M+)	This work
pZS*11-YFP	<i>pSC101* ori</i> , <i>Amp</i> <sup>R</sup> , <i>P</i> <sub><i>LtetO-1</i></sub> - <i>yfp</i>	Lab collection
pZA32-GFP	<i>p15A ori</i> , <i>Cam</i> <sup>R</sup> , <i>P</i> <sub><i>LlacO-1</i></sub> -GFP	Lab collection
pZS*11-R. <i>EcoRI</i>	<i>pSC101* ori</i> , <i>Amp</i> <sup>R</sup> , <i>P</i> <sub><i>LtetO-1</i></sub> -R. <i>EcoRI</i>	This work
pZA32-M. <i>EcoRI</i>	<i>p15A ori</i> , <i>Cam</i> <sup>R</sup> , <i>P</i> <sub><i>LlacO-1</i></sub> -M. <i>EcoRI</i>	This work
pZA32-M. <i>EcoRV</i>	<i>p15A ori</i> , <i>Cam</i> <sup>R</sup> , <i>P</i> <sub><i>LlacO-1</i></sub> -M. <i>EcoRI</i>	This work
$\lambda$ <i>vir</i>	Virulent mutant of Phage $\lambda$	Allan Campbell

### 3.4.9 Strain and plasmid construction

The  $P_{sulA}$  promoter region was PCR-amplified from the MG1655 chromosome using the 26\_5 and 27\_3 primers. This amplifies 240 bp long region (120 bp upstream and 120 bp downstream of the *sulA* start codon) and introduces an in-frame stop codon. The amplified fragment was digested with BamHI and XhoI and cloned into the pAH68-*frt-cam* plasmid. This is a CRIM-based plasmid (Haldimann & Wanner, 2001) previously constructed in our laboratory that integrates into the HK022 attachment site. The plasmid contains promoterless YFP and a chloramphenicol resistance marker flanked by FRT sites so it can be removed using the pCP20 plasmid (Doublet et al., 2008). DH5 $\alpha$   $\lambda$ *pir+* was used as host for cloning the plasmid. The resulting plasmid (**pMP017**) containing the  $P_{sulA}$ -*yfp* fusion was integrated into the MG1655 chromosome using the pAH69 helper plasmid. The Cm<sup>R</sup> marker was removed with pCP20 (**MP060**). The activity of the  $P_{sulA}$ -*yfp* reporter increased 40-fold in response to 1  $\mu$ g/ml mitomycinC (not shown). **MP062** was constructed by transduction (P1) of the  $\Delta att\lambda::(P_{N25}$ -tetR,  $P_{lacI^q}$ -*lacI*,  $Sp^R$ ) cassette from Frag1B into MG1655, followed transduction of the  $\Delta attHK022::(P_{sulA}$ -*yfp*, Cm<sup>R</sup>-*frt*) allele and removal of the Cm<sup>R</sup> marker (pCP20). **MP066** was constructed by P1 transduction of the  $\Delta attP21::(P_R$ -*mCherry*, Cm<sup>R</sup>-*frt*) allele into MP060 followed by removal of the Cm<sup>R</sup> marker (pCP20). All the  $\Delta recA$  strains were constructed by recombineering (Thomason, Sawitzke, Li, Costantino, & Court, 2014) of the  $\Delta recA$ , *kan*<sup>R</sup>-*frt* deletion PCR product into either MG1655 (**MP070**), MP060 (**MP064**) or MP066 (**MP068**) followed by removal of the *kan*<sup>R</sup> marker (pCP20). The  $\Delta recA$ , *kan*<sup>R</sup>-*frt* PCR product was obtained in a PCR reaction using the 5delRecA and 3delRecA primers and the pKD13 plasmid (Cox et al., 2007) as the template. Genotypes of all the strains were verified by PCR. The **MG1655**  $P_R$ -*venus* and **MG1655**  $P_R$ -*cerulean* were constructed by integrating the  $P_R$ -*venus/cerulean* cassette into the P21 attachment site using a CRIM-based plasmid with removable chloramphenicol resistance marker (the markers were removed from the host strain using pCP20).

To prevent transcription from the  $P_{tet}$  promoter of pBR322, which could affect the expression levels of the cloned RM systems, **BR322 $\Delta$ P<sub>tet</sub>** was constructed by double-digestion of pBR322 with BamHI and HindIII, blunting the ends using the T4 polymerase and relegation. The EcoRI (R+M+) and EcoRI (R-M+) genes were PCR amplified from pIK166 and pIK167 respectively using the 95\_5 and 103\_3 primers, followed by digestion by XhoI and

HindIII (partial digestion of the EcoRI (R+M+) fragment was necessary, as the R.EcoRI gene contains additional HindIII site). The fragments were cloned into pBR322 via the HindIII and Sall restriction sites, giving rise to **pBR322 $\Delta$ P<sub>tet</sub>EcoRI (R+M+)** and **pBR322 $\Delta$ P<sub>tet</sub>EcoRI (R-M+)**. Analogously, The EcoRV (R+M+) and EcoRV (R-M+) fragments were amplified from pYNEC107 and pYNEC117 respectively using the 96\_5 and 99\_3 primers, followed by digestion with HindIII and XhoI and cloning into the pBR322 backbone using the HindIII and Sall restriction sites, resulting in **pBR322 $\Delta$ P<sub>tet</sub>EcoRV (R+M+)** and **pBR322 $\Delta$ P<sub>tet</sub>EcoRV (R-M+)**. **pZS\*11-R.EcoRI** was constructed by amplifying the R.EcoRI coding sequence from pIK166 using the 87\_5 and 88\_3 primers. The PCR product was digested by KpnI and XbaI and cloned into pZS\*11-Venus using the same restriction sites. The M.EcoRI coding sequence was amplified from pIK166 using the 89\_5 and 90\_3 primers. The resulting fragment was digested with KpnI and XbaI and cloned into pZS\*11-Venus. From there, the M.EcoRI-T1 fragment (M.EcoRI + terminator sequence) was cleaved by KpnI and AvrII and cloned into pZA32-GFP using the same restriction sites, resulting in **pZA\*32-M.EcoRI**. The M.EcoRV coding sequence was amplified from pYNEC107 using the 93\_5 and 94\_3 primers. The resulting fragment was digested with KpnI and XbaI and cloned into pZS\*11-Venus. From there, the M.EcoRV-T1 fragment (M.EcoRV + terminator sequence) was cleaved by KpnI and AvrII and cloned into pZA32-GFP using the same restriction sites, resulting in **pZA\*32-M.EcoRV**. MP062 was used as the host for cloning of the inducible plasmids. The identity of all the plasmids was verified by sequencing of the cloned regions. All reagents were from Sigma. All enzymes were from NEB. Phusion® High-Fidelity DNA Polymerase (NEB) was used to amplify fragments for cloning. All the cloned fragments were verified by PCR.

### 3.4.10 List of primers

Name	Sequence (5' to 3') <sup>a</sup>
26_5	aattaaCTCGAGTTATGTTTTCCCGTCACCAA
27_3	aattaaGGATCCTTCGCGATAGACAACCTTCAC
5delRecA	TGACTATCCGGTATTACCCGGCATGACAGGAGTAAAAATGGGGGATCCGTC <b>GACCTGCAGTT</b>
3delRecA	AAGGGCCGCAGATGCGACCCTTGTGTATCAAACAAGACGATGTAGGCTGG <b>AGCTGCTTC</b>
87_5	aaGGTACCCATGTCTAATAAAAAACAGTC
88_3	aaTCTAGATCACTTAGATGTAAGCTG
89_5	aaGGTACCATGGCTAGAAATGCAAC
90_3	aaTCTAGACTAACAACATCAATTACTTTTG
93_5	AAggtaccGGTTCTTATTATGAAAGATAAAG
94_5	AATCTAGACCATCACTCTTCAATTACC
95_5	aatCTCGAGCATCTGGTTGCATAGGTAT
96_3	aatCTCGAGGGAATAGTCTGATGCTAAAT
99_3	aaAAGCTTCCATCACTCTTCAATTACC
103_3	aaAAGCTTCTAACAACATCAATTACTTTTG

<sup>a</sup> Annealing sequences are shown in bold

## **4 Phage-host population dynamics promotes prophage acquisition in bacteria with innate immunity**

This chapter contains results and their discussion published in: Pleška M, Lang M, Refardt D, Levin BR, Guet CC (2018) Phage-host population dynamics promotes prophage acquisition in bacteria with innate immunity. *Nature Ecology & Evolution* (in press). The chapter was written in collaboration with Moritz Lang (IST Austria), who constructed and analyzed the mathematical models.

### **4.1 Summary**

Temperate bacteriophages integrate in bacterial genomes as prophages and represent an important source of genetic variation for bacterial evolution, frequently transmitting fitness-augmenting genes such as toxins responsible for virulence of major pathogens. However, only a fraction of phage infections are lysogenic and lead to prophage acquisition, whereas the majority are lytic and kill the infected bacteria. Unless able to discriminate lytic from lysogenic infections, mechanisms of immunity to bacteriophages are expected to act as a double-edged sword and increase the odds of survival at the cost of depriving bacteria of potentially beneficial prophages. We show that although restriction-modification systems as mechanisms of innate immunity prevent both lytic and lysogenic infections indiscriminately in individual bacteria, they increase the number of prophage-acquiring individuals at the population level. We find that this counterintuitive result is a consequence of phage-host population dynamics, in which restriction-modification systems delay infection onset until bacteria reach densities at which the probability of lysogeny increases. These results underscore the importance of population-level dynamics as a key factor modulating costs and benefits of immunity to temperate bacteriophages.

## 4.2 Introduction

Restriction-modification (RM) systems abound in the Archaea as well as the Bacteria, with multiple RM systems frequently residing in a single genome (Oliveira et al., 2014; Vasu & Nagaraja, 2013; Wilson & Murray, 1991). Typically composed of two enzymatic activities, RM systems represent a minimal mechanism of self- / non-self discrimination (Murray, 2002). This essential biological role is realized by the restriction endonuclease, which cleaves exogenous “non-self” DNA at well-defined sequences termed restriction sites, while the cognate methyltransferase prevents cleavage of endogenous “self” restriction sites by methylation. As a result, RM systems provide bacteria with innate immunity (Abedon, 2012) against bacteriophage (phage) infections (Tock & Dryden, 2005). While this function has typically been investigated in the context of obligatorily lytic phages, the nature of interactions between RM systems and temperate phages remains elusive.

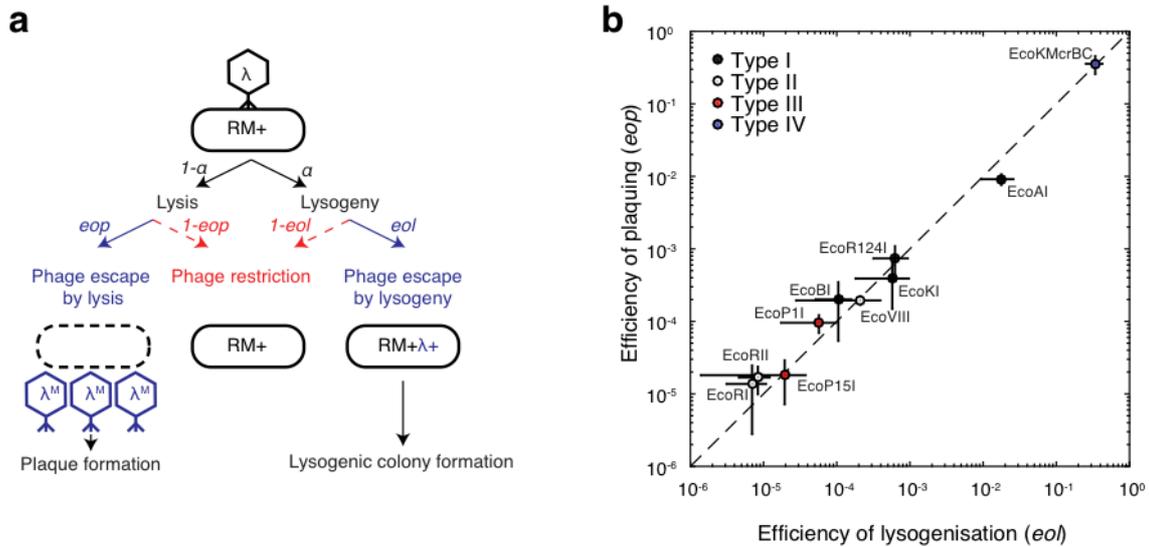
In addition to horizontal spread by lysis, temperate phages can enter bacterial genomes as prophages and transmit vertically in the process termed lysogeny (Lwoff, 1953). Prophages are a prevalent feature of bacterial genomes (Bobay, Rocha, & Touchon, 2013) and can constitute as much as 20% of their size (Casjens, 2003). A large number of prophages carry genes that increase fitness of their bacterial hosts (Edlin, Lin, & Bitner, 1977; Lin, Bitner, & Edlin, 1977; Obeng, Pratama, & Elsas, 2016; Oliver, Degnan, Hunter, & Moran, 2009), examples of which include determinants of microbial pathogenicity and virulence (Barondess & Beckwith, 1990; Brüssow, Canchaya, Hardt, & Bru, 2004; O’Brien et al., 1984; Waldor & Mekalanos, 1996), genes increasing resistance to adverse environments (X. Wang et al., 2010), as well as those controlling biofilm formation (Rice et al., 2009). Moreover, prophages can confer immunity to phage superinfection (Bondy-Denomy et al., 2016), serve as allopathic agents during invasion of new environments (Brown, Le Chat, De Paepe, & Taddei, 2006), and cause beneficial mutations (Davies et al., 2016). An infection by a temperate phage can thus result either in host death or acquisition of a potentially beneficial prophage. It has recently been shown that some CRISPR/Cas systems, a type of bacterial adaptive immunity (Barrangou et al., 2007), tolerate lysogenic infections and only interfere with lysis (Goldberg, Jiang, Bikard, & Marraffini, 2014), thus protecting their hosts without compromising prophage acquisition. However, tolerance to lysogeny is not a

property of all CRISPR/Cas types (Edgar & Qimron, 2010) and whether it occurs in the context of RM systems is not known. Can RM systems as mechanisms of innate immunity distinguish between lytic and lysogenic infections, or do they act as a barrier to prophage acquisition?

## 4.3 Results

### 4.3.1 RM systems prevent lytic and lysogenic infections indiscriminately at the level of individual bacteria

To address this question, we first examined eleven RM systems originally isolated from *Escherichia coli* for their ability to prevent lytic and lysogenic infections. The RM systems we tested represented all four types into which RM systems are classified based on their molecular composition (R. Roberts, 2003). Our particular concern was whether, for any of these systems, the probability of a temperate phage escaping restriction during establishment of lysogeny substantially exceeds the probability of escape during lysis. As measures of these probabilities, we used the efficiency of lysogen formation (*eol*) and the efficiency of plaque formation (*eop*) (**Figure 4-1A**), respectively. The temperate  $\lambda$  *kan* phage used in these and subsequent experiments carried a gene rendering lysogens resistant to kanamycin, which allows for their direct selection. For ten of the eleven tested RM systems, *eop* and *eol* did not differ significantly (**Figure 4-1B** and **Table 4-1**). The *eop* of the eleventh RM system, EcoRV, was the lowest observed ( $1.96 \cdot 10^{-8} \pm 1.53 \cdot 10^{-8}$ ) and because its *eol* was below the detection limit of the assay ( $< 10^{-7}$ ) the two efficiencies could not be directly compared. However, the fact that no lysogenic colonies were obtained suggested that *eol* of EcoRV was not substantially higher than its *eop*. We interpret these results to be inconsistent with the hypothesis that any of the tested RM systems possess a molecular mechanism that would allow individual bacteria to tolerate lysogenic infections. In other words, RM systems cleave DNA of phages entering the lytic and lysogenic pathway indiscriminately, and represent a barrier to prophage acquisition in individual bacteria.



#### Figure 4-1: RM systems represent a barrier to prophage acquisition in individual bacteria

**(A)** Possible outcomes following an infection of a RM-carrying bacterium by unmethylated phage  $\lambda$ . The phage enters the lysogenic pathway with the probability of lysogeny  $\alpha$  and the lytic pathway with the probability  $1 - \alpha$ . The RM system fails to restrict lytic and lysogenic infections with probabilities  $eop$  and  $eol$ , respectively (solid blue arrows). Phage escape during lytic infection leads to spread of methylated phage and formation of a plaque. Phage escape during lysogenic infection leads to integration of the phage DNA into the genome of the host and formation of a lysogen, giving rise to a lysogenic colony. Infections in which the phage does not escape lead to phage restriction (dashed red arrows). The  $eop$  is defined as the number of plaque forming units ( $pfu$ ) obtained on lawns of bacteria carrying a RM system (RM+), relative to the total  $pfu$  obtained on lawns of RM- bacteria ( $eop = pfu_{RM+}/pfu_{RM-}$ ). Analogously, the  $eol$  is defined as the relative number of lysogenic colony-forming units ( $lcfu$ ) obtained after infection of RM+ and RM- bacteria at the phage/bacteria ratio of 1 ( $eol = lcfu_{RM+}/lcfu_{RM-}$ ).

**(B)** RM systems restrict phages entering lytic and lysogenic pathways indiscriminately. Means of six replicates from two sets of experiments (three independent biological replicates per experiment) are shown. Error bars represent standard deviations. P values were calculated by unpaired, one-sided Welch's t-test ( $H_0: eol \leq eop$ ). The results were not significant after Bonferroni correction for multiple comparisons (**Table 4-1**). The dashed line corresponds to  $eop = eol$ . The limit of detections is approximately  $10^{-10}$  and  $10^{-7}$  for  $eop$  and  $eol$ , respectively. EcoRV (Type IIP) is not plotted since its  $eol$  was below the detection limit of the assay. The corresponding  $eop$  was  $1.9 \cdot 10^{-8}$ .

**Table 4-1: eop and eol of bacteria carrying different RM systems**

RM System	Type	eop <sup>a</sup>		eol <sup>a</sup>		P value <sup>b</sup>
		Mean	SD	Mean	SD	
EcoAI	IB	$9.14 \cdot 10^{-3}$	$1.53 \cdot 10^{-3}$	$1.77 \cdot 10^{-2}$	$8.17 \cdot 10^{-3}$	0.03
EcoBI	IA	$7.38 \cdot 10^{-4}$	$3.50 \cdot 10^{-4}$	$6.26 \cdot 10^{-4}$	$3.07 \cdot 10^{-4}$	0.91
EcoKI	IA	$3.90 \cdot 10^{-4}$	$2.41 \cdot 10^{-4}$	$5.74 \cdot 10^{-4}$	$3.93 \cdot 10^{-4}$	0.18
EcoR124I	IC	$2.01 \cdot 10^{-4}$	$1.47 \cdot 10^{-4}$	$1.05 \cdot 10^{-4}$	$5.30 \cdot 10^{-5}$	0.72
EcoRI	IIP	$1.37 \cdot 10^{-5}$	$1.09 \cdot 10^{-5}$	$6.97 \cdot 10^{-6}$	$3.83 \cdot 10^{-6}$	0.90
EcoRII	IIEP	$1.68 \cdot 10^{-5}$	$6.87 \cdot 10^{-6}$	$8.35 \cdot 10^{-6}$	$3.72 \cdot 10^{-6}$	0.85
EcoRV	IIP	$1.95 \cdot 10^{-8}$	$1.53 \cdot 10^{-8}$			
EcoVIII	IIP	$1.94 \cdot 10^{-4}$	$2.38 \cdot 10^{-5}$	$2.11 \cdot 10^{-4}$	$1.83 \cdot 10^{-4}$	0.98
EcoP1	III	$9.57 \cdot 10^{-5}$	$2.62 \cdot 10^{-5}$	$5.71 \cdot 10^{-5}$	$3.97 \cdot 10^{-5}$	0.96
EcoP15I	III	$1.81 \cdot 10^{-5}$	$1.09 \cdot 10^{-5}$	$1.94 \cdot 10^{-5}$	$1.80 \cdot 10^{-5}$	0.44
EcoKMcrBC	IV	$3.57 \cdot 10^{-1}$	$9.83 \cdot 10^{-2}$	$3.40 \cdot 10^{-1}$	$8.51 \cdot 10^{-2}$	0.62

<sup>a</sup> Calculated from six replicates in two sets of experiments (three independent biological replicates per experiment).

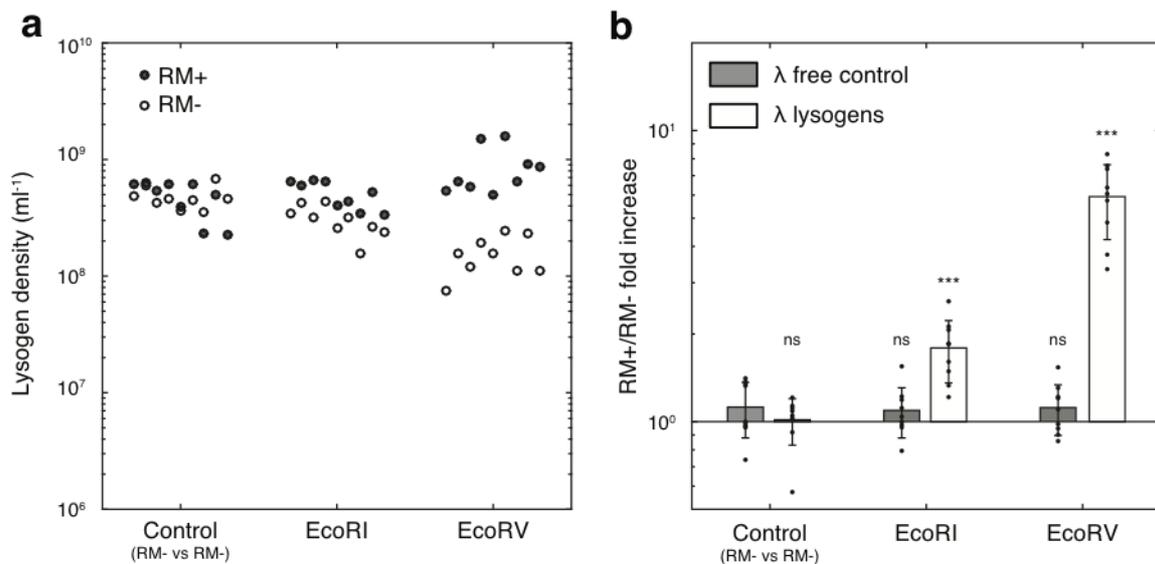
<sup>b</sup> Calculated by unpaired, one-sided Welch's t-test ( $H_0: eol \leq eop$ ). P values were not significant after Bonferroni correction for multiple comparisons  $\alpha^* = \frac{0.05}{10} = 0.005$ .

### 4.3.2 RM systems promote lysogenic infections at the population level

To understand how the RM systems affect prophage acquisition at the population level, we infected mixed RM+ and RM- cultures growing in a defined minimal medium (M9 + 0.4% maltose) with  $\lambda$  *kan* and measured lysogen densities after 24 hours of incubation. RM+ bacteria carried either EcoRI or EcoRV, two RM systems chosen due to their widely different probabilities of phage escape ( $eop = 1.37 \cdot 10^{-5}$  and  $1.96 \cdot 10^{-8}$  for EcoRI and EcoRV, respectively (**Figure 4-1B**)). RM+ and RM- bacteria were isogenic, except for the chromosomal markers (*ara<sup>-</sup> cat<sup>+</sup>* vs. *ara<sup>+</sup> cat<sup>-</sup>* for RM+ vs. RM- in all experiments). Contrary to what we anticipated based on RM systems acting as a barrier to prophage acquisition in individual bacteria, RM+ bacteria formed more lysogens as compared to RM- (**Figure 4-2A**) and were therefore more likely to acquire the prophage at the population level. Indeed, the RM+/RM- ratio of lysogens 24 hours post infection significantly exceeded the initial RM+/RM- ratio in experiments with both EcoRI as well as EcoRV, whereas no significant change was observed in phage free controls (**Figure 4-2B**). Notably, the RM+/RM- fold increase following infection and subsequent selection for lysogens was more pronounced for EcoRV, which cleaves  $\lambda$  *kan* with a higher efficiency and was thus expected to act as a stronger barrier to prophage acquisition. We obtained consistent results in a medium with a different composition (M63 + 0.4% maltose) (**Figure 4-3**). Increase in the RM+/RM- ratio also occurred when we measured the ratio of total bacteria without selection for lysogens in a series of daily transfers (**Figure 4-4**). In these experiments both the RM+ and RM- populations were dominated by lysogens at the time of the first transfer and remained such until the end of the experiment. Consistent with results of a previous study (Pleška et al., 2016), we observed a small fitness cost of EcoRI, but not EcoRV, in the absence of the phage (**Figure 4-4**).

The above experiments show that although RM systems do not allow individual bacteria to selectively tolerate lysogenic infections, they can increase the number of prophage acquiring bacteria at the population level. In a series of analogous experiments, we tested how this ability depends on the initial population composition by varying the initial density of phages and bacteria, as well as the initial RM+/RM- ratio. Upon infection by  $\lambda$  *kan* and selection for lysogeny, bacteria carrying EcoRI and EcoRV produced more lysogens as

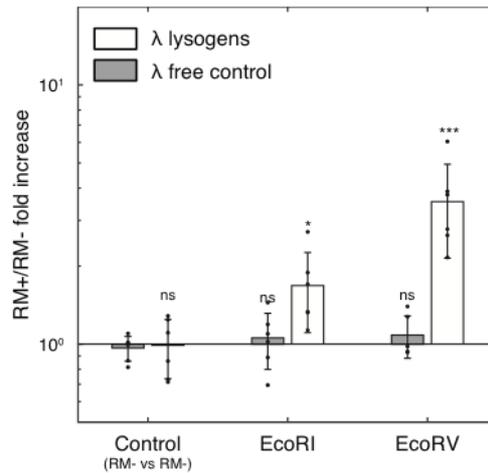
compared to RM<sup>-</sup> bacteria under a wide range of initial conditions (**Figure 4-5**). In experiments with EcoRI, the effect increased at low initial phage densities, low initial bacterial densities, as well as low RM<sup>+</sup> to RM<sup>-</sup> ratios, where it reached the levels observed for EvoRV. In experiments with EcoRV on the other hand, the fold increase in RM<sup>+</sup>/RM<sup>-</sup> ratio was quantitatively consistent across most conditions tested, only decreasing at very high initial bacterial densities. Furthermore, in experiments with EcoRV at the initial RM<sup>+</sup>/RM<sup>-</sup> ratio of 100:1, no RM<sup>+</sup> lysogens were detected and RM<sup>-</sup> bacteria formed only very few lysogens ( $10^2$  RM<sup>-</sup> lysogens/ml at the initial ratio of 100:1 vs.  $10^7$  RM<sup>-</sup> lysogens/ml at the initial ratio of 10:1). This is likely because the highly abundant population of immune bacteria restricts the majority of phages and thus protects the sensitive subpopulation from infection. The results show that unless overly abundant, RM systems can promote lysogen formation on a population level under a broad range of initial conditions.



**Figure 4-2: RM systems promote prophage acquisition at the population level**

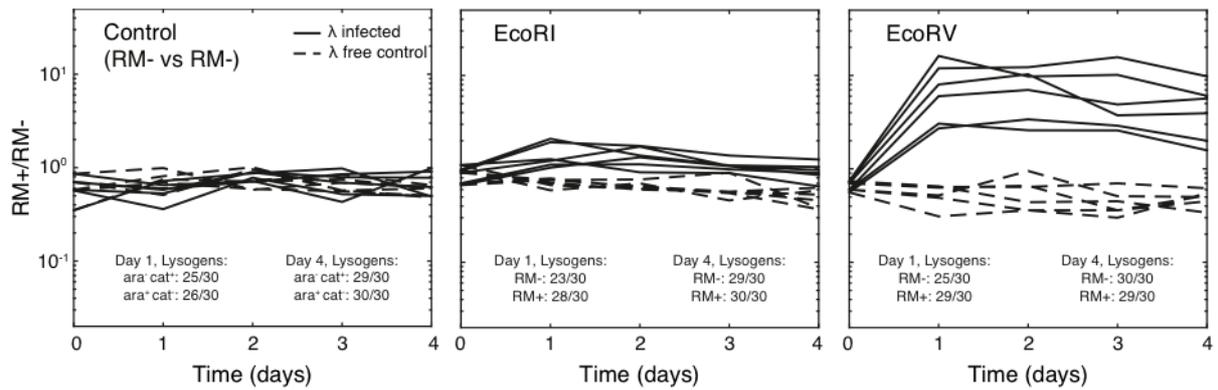
**(A)** Densities of RM+ (filled circles) and RM- (open circles) lysogens 24 hours after infection by  $\lambda$  *kan*. Initially, the cultures contained  $10^7$  bacteria/ml,  $10^5$  phages/ml, and the RM+/RM- ratio of approximately 1. 5/5 RM+ and 5/5 RM- colonies in each experiment released free phage and were thus lysogenic. 5/5 RM+ colonies in each experiment retained restriction activity as tested by infection with unmodified  $\lambda$  *vir* phage. None of the tested colonies was envelope resistant as tested by  $\lambda$  *vir* phage modified by the respective RM system. Nine replicates from three sets of experiments (three independent biological replicates per experiment) are shown.

**(B)** Gray bars represent RM+/RM- ratios of non-lysogens after 24-hour incubation in the absence of the phage. White bars represent ratios of RM+/RM- lysogens 24 hours after infection (calculated from data shown in **Figure 4-2A**). All measurements were normalized by the respective initial RM+/RM- ratio. For the control (RM- vs. RM-) experiments, the y-axis depicts the *ara<sup>-</sup> cat<sup>+</sup>/ara<sup>+</sup> cat<sup>-</sup>* ratio. Nine replicates from three sets of experiments (three independent biological replicates per experiment) are shown as individual data points. Corresponding means are shown as bars. Error bars represent standard deviations. P values were calculated by multiple linear regression with interaction terms, with the logarithm of RM+/RM- fold increase as a continuous dependent variable and strain identity (EcoRI/EcoRV/Control) and treatment ( $\lambda$  present/ $\lambda$  absent) as categorical independent variables. Control (RM- vs. RM-) was set as intercept. Reported values are the p values for the t statistics of the interaction terms. Asterisks indicate the level of significance (ns = not significant).



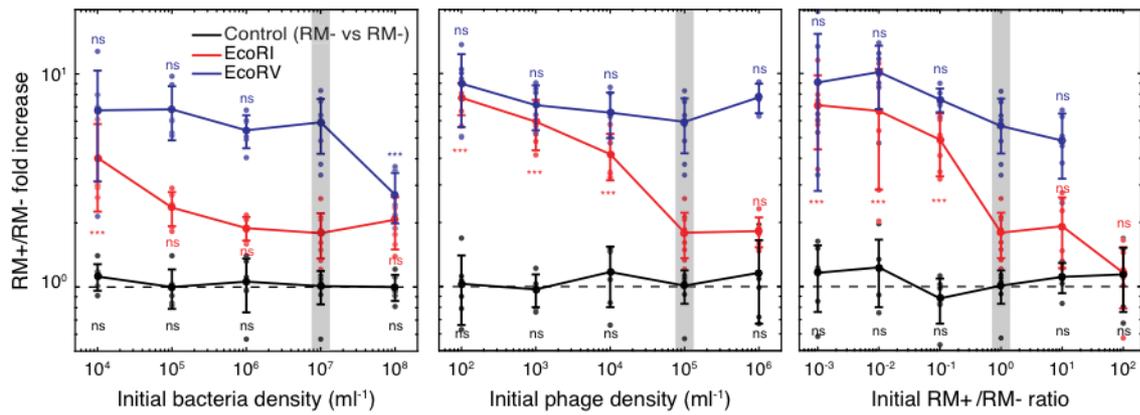
**Figure 4-3: Prophage acquisition in M63 medium**

Black bars represent RM+/RM- ratios of non-lysogens after 24-hour incubation in the absence of the phage. White bars represent ratios of RM+/RM- lysogens 24 hours after infection. All measurements were normalized by the respective initial RM+/RM- ratio. For the control (RM- vs. RM-) experiments, the y-axis depicts the *ara<sup>-</sup>cat<sup>+</sup>/ara<sup>+</sup>cat<sup>-</sup>* ratio. Six replicates from three sets of experiments (three independent biological replicates per experiment) are shown as individual data points. Corresponding means are shown as bars. Error bars represent standard deviations. P values were calculated by multiple linear regression with interaction terms, with the logarithm of RM+/RM- fold increase as a continuous dependent variable and strain identity (EcoRI/EcoRV/Control) and treatment ( $\lambda$  present/ $\lambda$  absent) as categorical independent variables. Control (RM- vs. RM-) was set as intercept. Reported values are the p values for the t statistics of the interaction terms. Asterisks indicate the level of significance (ns = not significant).



**Figure 4-4: Prophage acquisition in serially transferred cultures**

The plots show RM+/RM- ratios of total bacteria in control experiments (left) and experiments with EcoRI (middle) and EcoRV (right). Solid lines represent cultures infected with unmethylated  $\lambda$  *kan* at day 0. Dashed lines represent phage-free cultures. Constant phage density of approximately 10<sup>7</sup> pfu/ml was maintained in all infected cultures throughout all experiments (not shown). After the first and fourth day of the experiment, we tested five RM+ and RM- colonies from each of the six infected cultures for lysogeny. The numbers of lysogenic colonies are shown at the bottom of each figure. Six independent replicates are shown. Average selection coefficients (Dykhuizen & Hartl, 1983) (n=6) calculated from the phage free experiments were  $-0.02 \pm 0.07$  (Control),  $-0.14 \pm 0.05$  (EcoRI), and  $0.08 \pm 0.07$  (EcoRV) (mean  $\pm$  SD, day<sup>-1</sup>). The selection coefficient due to EcoRI was significantly different from the control (P = 0.006), whereas the selection coefficient due to EcoRV was not (p = 0.12). P values were calculated by linear regression, with the selection coefficient as a continuous dependent variable and strain identity as a categorical independent variable, comparing individual strains to the control (RM- vs. RM-).



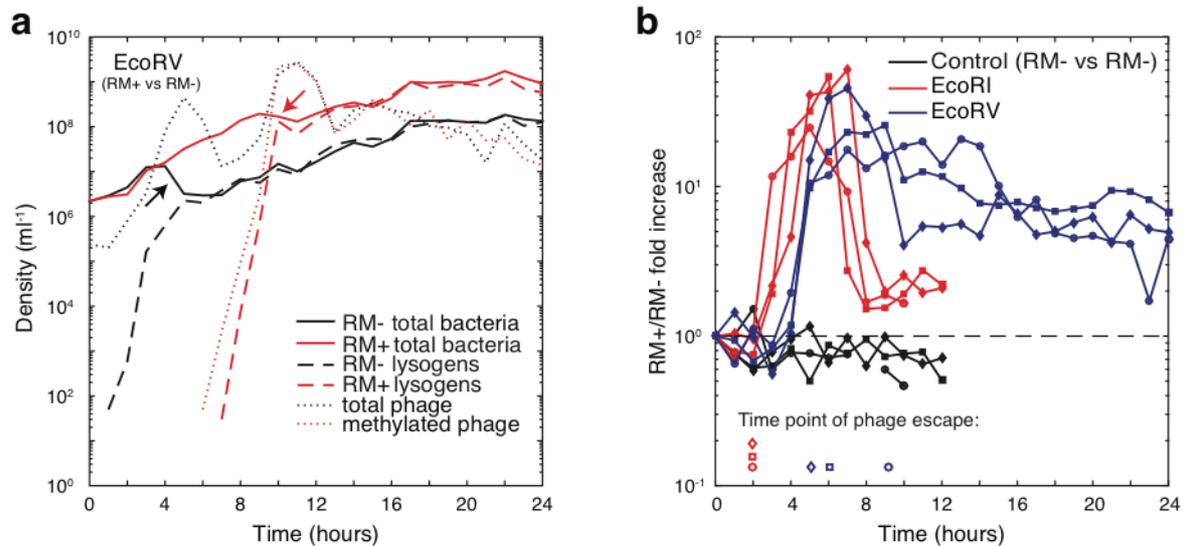
**Figure 4-5: RM systems promote prophage acquisition under a wide range of initial conditions**

The plots show the effect of initial bacteria density (left), phage density (middle) and the initial RM+ to RM- ratio (right) on the ability of RM systems to promote lysogeny. Grey regions correspond to conditions examined in **Figure 4-2A**. Each point represents the ratio of RM+ to RM- lysogens 24 hours after infection normalized by the respective initial RM+ / RM- ratio. For the control (RM- vs. RM-) experiments, the y-axis depicts the  $ara^- cat^+ / ara^+ cat^-$  ratio. Means of six replicates from two sets of experiments (three independent biological replicates per experiment) are shown, except for the points in grey regions, which represent means of nine replicates from three experiments. Small points represent results of individual replicate experiments, large points represent their means. Error bars represent standard deviations. P values were calculated by multiple linear regression with interaction terms, with the logarithm of RM+ / RM- fold increase as a continuous dependent variable and strain identity (EcoRI/EcoRV/Control) and treatment (initial density in left and middle plot, initial ratio in right plot) as categorical independent variables. A single linear regression model was fit to data presented in each panel with values in grey regions as intercepts. Reported values are the p values for the t statistics of the interaction terms and indicate values significantly different from those shown in grey regions. Asterisks indicate levels of significance (ns = not significant).

### 4.3.3 RM systems postpone the onset of infection

To elucidate the mechanisms responsible for the unexpectedly high number of prophage-acquiring RM<sup>+</sup> bacteria, we followed the population dynamics by estimating phage and bacterial densities at one-hour intervals. In **Figure 4-6A**, we show representative results obtained in an experiment with EcoRV. During the initial four hours, the density of unmethylated phage increased as a fraction of RM<sup>-</sup> bacteria was killed by unmethylated phage (**Figure 4-6A**, black arrow). At the same time, a number of RM<sup>-</sup> bacteria acquired the prophage and formed lysogens. Since lysogens are immune to secondary infections, RM<sup>-</sup> lysogens survived and grew despite phage densities being still high. RM<sup>+</sup> bacteria resisted infection and grew exponentially until five hours into the experiment, when the density of unmethylated phage peaked and first methylated phage appeared as a result of phage escape, which marked a turning point in the experiment. Because methylated phages evade restriction, they rapidly multiplied on RM<sup>+</sup> bacteria and a second wave of infection ensued. All RM<sup>+</sup> bacteria were either lysed or lysogenized by ten hours into the experiment. Importantly, the drop in the density of RM<sup>+</sup> bacteria due to killing by methylated phage (**Figure 4-6A**, red arrow) was substantially smaller than the earlier drop in the density of RM<sup>-</sup> bacteria. As a result, the final density of RM<sup>+</sup> lysogens exceeded the density of RM<sup>-</sup> lysogens.

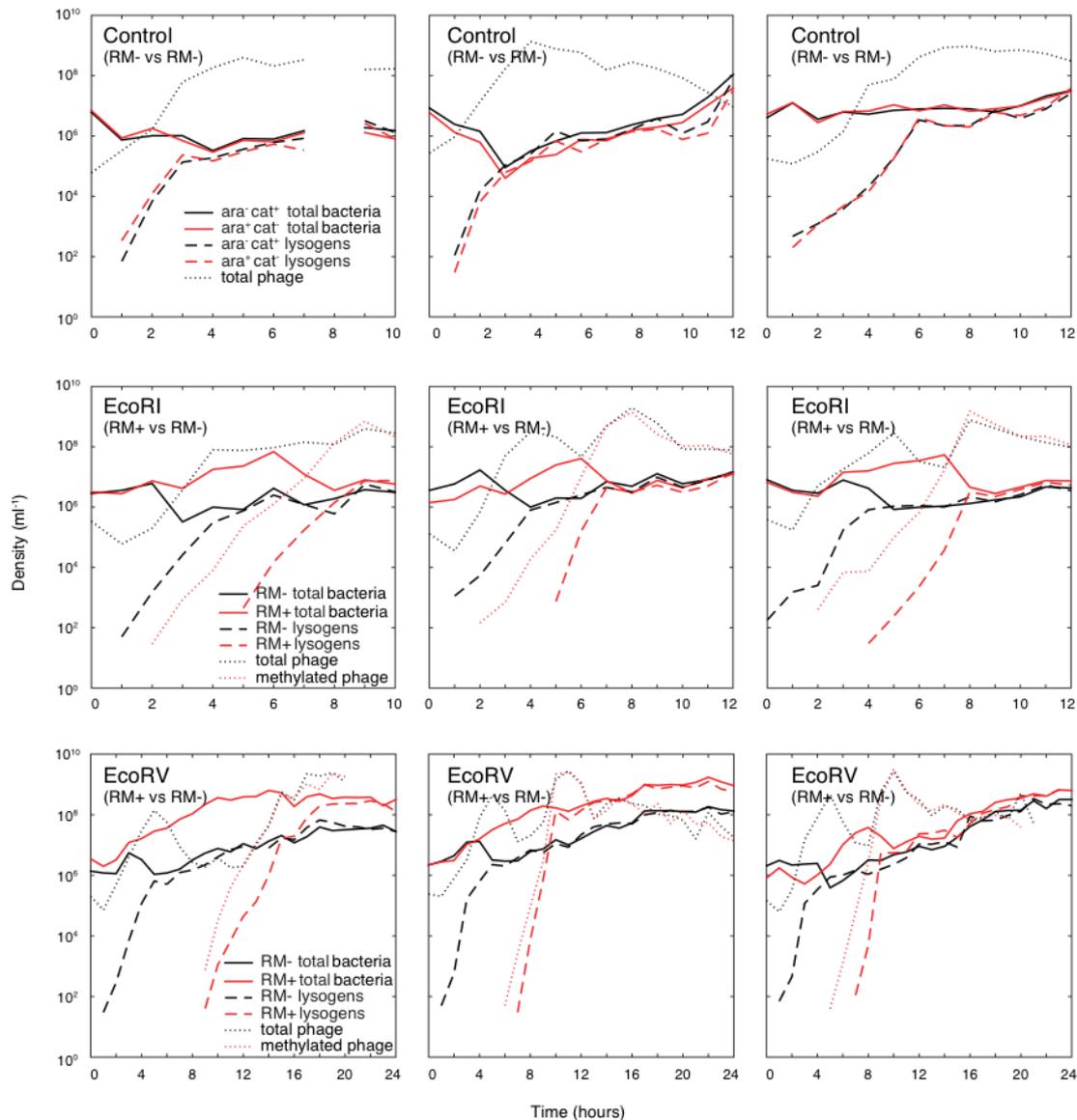
The difference in the fractions of RM<sup>+</sup> and RM<sup>-</sup> populations killed during the two subsequent waves of infection was apparent from observing the dynamics of the RM<sup>+</sup>/RM<sup>-</sup> ratio in three sets of replicate experiments (**Figure 4-6B**). In experiments with both RM systems, increases in the RM<sup>+</sup>/RM<sup>-</sup> ratio due to killing of RM<sup>-</sup> bacteria by unmethylated phage were more extensive than subsequent drops due to killing of RM<sup>+</sup> bacteria by methylated phage. However, there were also subtle differences between experiments with the two RM systems. In the case of EcoRV, phage escape of phage occurred later and was more variable in time between replicate experiments. Furthermore, killing of RM<sup>+</sup> bacteria occurred later and was smaller in magnitude in experiments with this RM system. We observed no significant ratio changes in experiments controlling for the effects of chromosomal markers (**Figure 4-6B**). RM systems thus did not prevent infection completely, but merely postponed its onset until methylated phage appeared and spread in the initially immune population.



**Figure 4-6: RM systems delay the onset of infection**

**(A)** Representative population dynamics in an experiment with EcoRV bacteria. The culture was infected with unmethylated  $\lambda$  *kan* at time 0. Phage and bacterial densities were measured at one-hour intervals. The limit of detection was 10 bacteria/phages per ml. The arrows correspond to killing of RM- (black) and RM+ (red) bacteria by unmethylated and methylated phage, respectively.

**(B)** Dynamics of RM+/RM- ratio in three sets of independent replicate experiments. All measurements were normalized by the respective initial RM+/RM- ratio. For the control (RM- vs. RM-) experiments, the y-axis depicts the  $ara^- cat^+ / ara^+ cat^-$  ratio. Open symbols at the bottom of the figure depict the times at which the first methylated phages were detected. The dynamics of control and EcoRI experiments were tracked for 12 hours, which was sufficient to fully capture the corresponding dynamics (**Figure 4-7**). Because methylated phage appear much later in experiments with EcoRV, the dynamics were tracked for 24 hours. All experiments were performed independently. Initial conditions corresponded to those used in **Figure 4-2**.



**Figure 4-7: Full dynamics of competition experiments in the presence of temperate phage**  
 Cultures were infected with unmethylated  $\lambda$  *kan* at time 0. Phage and bacterial densities were measured at one-hour intervals. The limit of detection was 10 bacteria (phages) / ml. For the control (RM- vs RM-) experiments, the strains differed only with respect to chromosomal markers. In the upper-left plot, the measurement at 8 hours is missing due to a mistake in the experimental procedure. In the second and third experiment with EcoRV, phage density at the last two time points was not assayed. All experiments were performed independently.

#### 4.3.4 Probability of lysogeny increases with population density

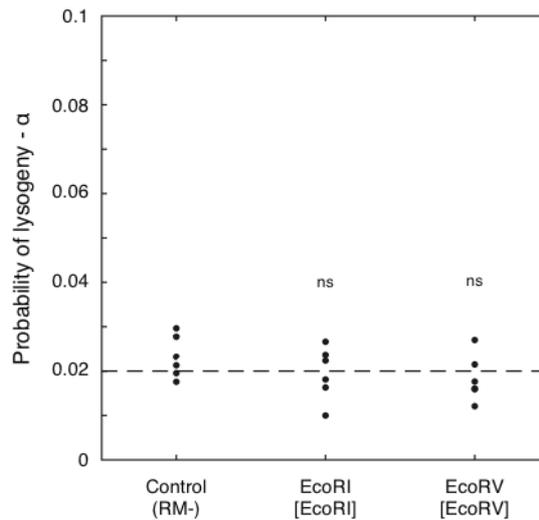
In addition, the above experiments suggested that the fraction of RM- and RM+ bacteria lysogenized during the two waves of infection were unequal. However, when measured in early exponential phase, presence of an RM system and phage methylation did not affect the probability of lysogeny, i.e. the probability that an infection by a phage, which does not get restricted, will result in lysogeny (**Figure 4-8**). Importantly, the population dynamics results (**Figure 4-6**) showed that RM systems substantially delayed the onset of infection, which could result in altered probability of lysogeny for bacteria infected at different growth phase. Indeed, the probability of lysogeny is known to depend on a variety of host-physiological parameters such as cell size (St-Pierre & Endy, 2008), or cAMP (Hong, Smith, & Ames, 1971) levels, and increases in stationary phase. Indeed, the probability of lysogeny under our experimental conditions increased over an order of magnitude as bacterial density increased (**Figure 4-9A**) and the two variables were strongly correlated (**Figure 4-9A** inlay).

We asked if this correlation can explain the increased number of prophage-acquiring RM+ bacteria by constructing and analyzing a mathematical model of population-level interactions between temperate phages and bacteria with RM systems (Material and Methods). Numerical solutions assuming density-dependent probability of lysogeny correctly predicted increased abundance of RM+ lysogens for both RM systems (**Figure 4-9B** and **Figure 4-10**). On the other hand, numerical solutions assuming a constant probability of lysogeny were inconsistent with the experimental results and predicted both the RM+ and RM- bacteria to produce equal number of lysogens. RM systems unable to selectively discriminate between lytic and lysogenic infections can thus benefit their hosts without compromising prophage acquisition simply by delaying infection onset. This delay allows the initially immune bacteria to reach densities at which the probability of an infection resulting in prophage acquisition is increased and the risk of lysis reduced.

Because more potent RM systems introduce longer delays in infection onset (**Figure 4-11**), the number of prophage-acquiring bacteria increases with decreasing probability of phage escape (**Figure 4-9B** inlay and **Figure 4-11**). In addition to predicting the increased probability of prophage acquisition by RM+ bacteria, our model thus also explained the

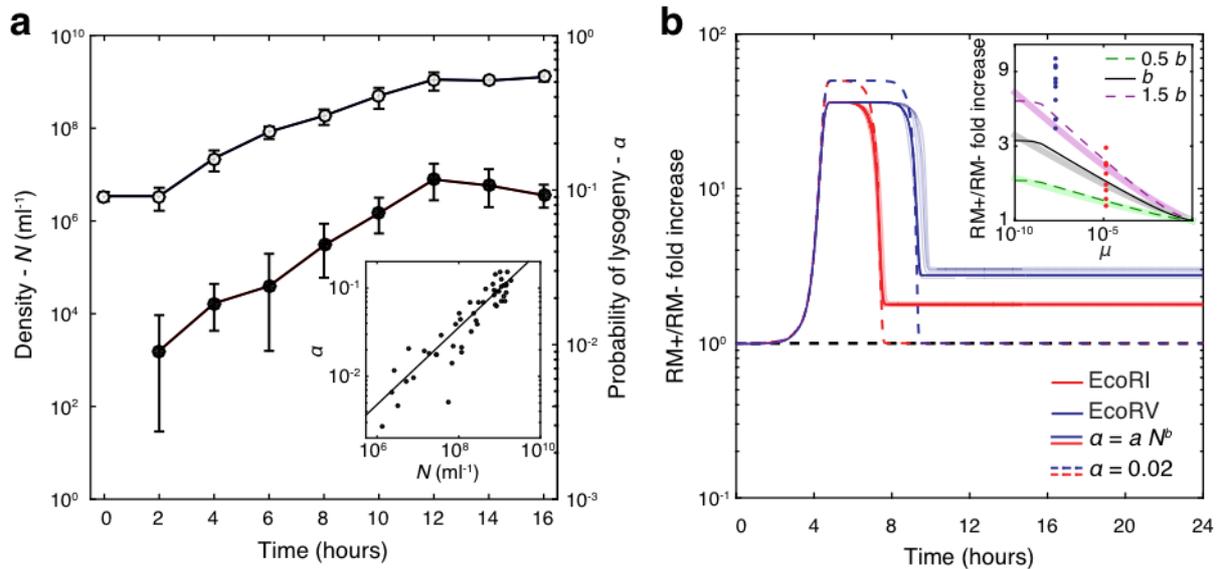
quantitative difference in effects caused by EcoRI and EcoRV. The model with density-dependent probability of lysogeny further predicted both RM systems to promote prophage acquisition under the wide range of initial conditions experimentally tested in **Figure 4-5 (Figure 4-12)**. While the model captured the general effect of the initial bacterial density and initial RM+/RM- ratio, it was inaccurate in predicting the increased effect at low initial phage densities observed for EcoRI. This disagreement is likely caused by simplifying assumptions used by the model, such as constancy of all parameters other than the probability of lysogeny, or not accounting for effects associated with multiple infections (Material and Methods).

Because phages escape restriction with a considerably low probability, we asked how stochastic effects could influence the model dynamics. Numerical simulations of a full stochastic version of the model yielded results quantitatively consistent with the deterministic model, demonstrating that stochastic “noise” played a relatively small role (**Figure 4-9B** and **Figure 4-13**). In contrast, the experimental results exhibited higher variability, which could be a result of additional sources of variation not captured by the stochastic model, such as small differences in initial conditions, sampling and measurement noise, as well as phenotypic variability. Importantly, both the deterministic and stochastic model predicted values slightly underestimating the experimentally measured results for EcoRV (**Figure 4-9B** inlay). This underestimation might be a result of parameter uncertainty. For example, in addition to the RM efficiency, the population dynamics critically depend on the rate at which the probability of lysogeny increases with cell density, and a moderate increase in this rate significantly improves the quantitative agreement between model predictions and the experimental data (**Figure 4-9B** inlay).



**Figure 4-8: RM systems do not alter prophage acquisition in individual bacteria**

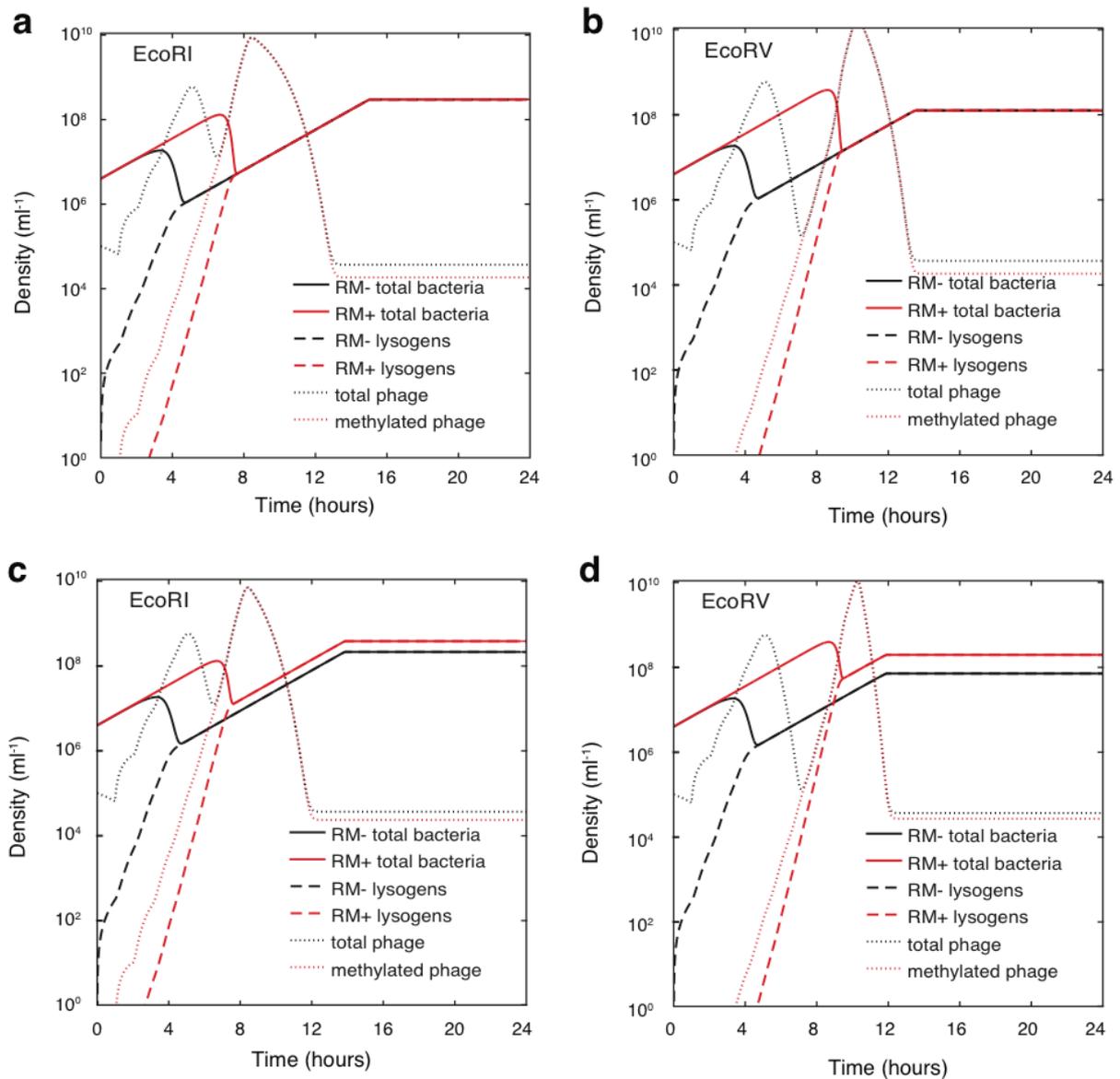
The phage used for infection was methylated by the respective RM system, also shown in square brackets (no modification in the control). The dashed line corresponds to the value of 2% estimated independently with *E. coli* MG1655 bacteria and used in the numerical simulations (**Table 4-2**). Infections were carried out at phage/bacteria ratio of 0.1. Six replicates from two sets of experiments (three independent biological replicates per experiment) are shown. EcoRI and EcoRV RM systems and the phage methylation state did not significantly affect the probability of lysogeny ( $P = 0.25$  and  $0.14$  for EcoRI and EcoRV, respectively).  $P$  values were calculated by linear regression, with the probability of lysogeny as a continuous dependent variable and strain identity (EcoRI/EcoRV/Control) as a categorical independent variable. Control (RM- vs. RM-) was set as intercept (ns = not significant).



**Figure 4-9: Delay in the onset of infection increases prophage acquisition**

**(A)** Bacterial density  $N$  (empty circles, left axis) and probability of lysogeny  $\alpha$  (full circles, right axis) measured in batch cultures sampled at two-hour intervals. Means of six replicates from two sets of experiments (three independent biological replicates per experiment) are shown. Error bars represent standard deviations. In all measurements, phage density was adjusted to reach phage/bacteria ratio of 0.1. The inlay shows  $\alpha$  as a function of bacterial density. Each point represents a single measurement averaged in the main figure. The solid line represents the power function  $\alpha(N) = a \cdot N^b$  fitted to the data, with  $a = 1.141 \cdot 10^{-5}$  and  $b = 0.4371$  (Pearson correlation = 0.92, P value =  $2.24 \cdot 10^{-20}$ ).

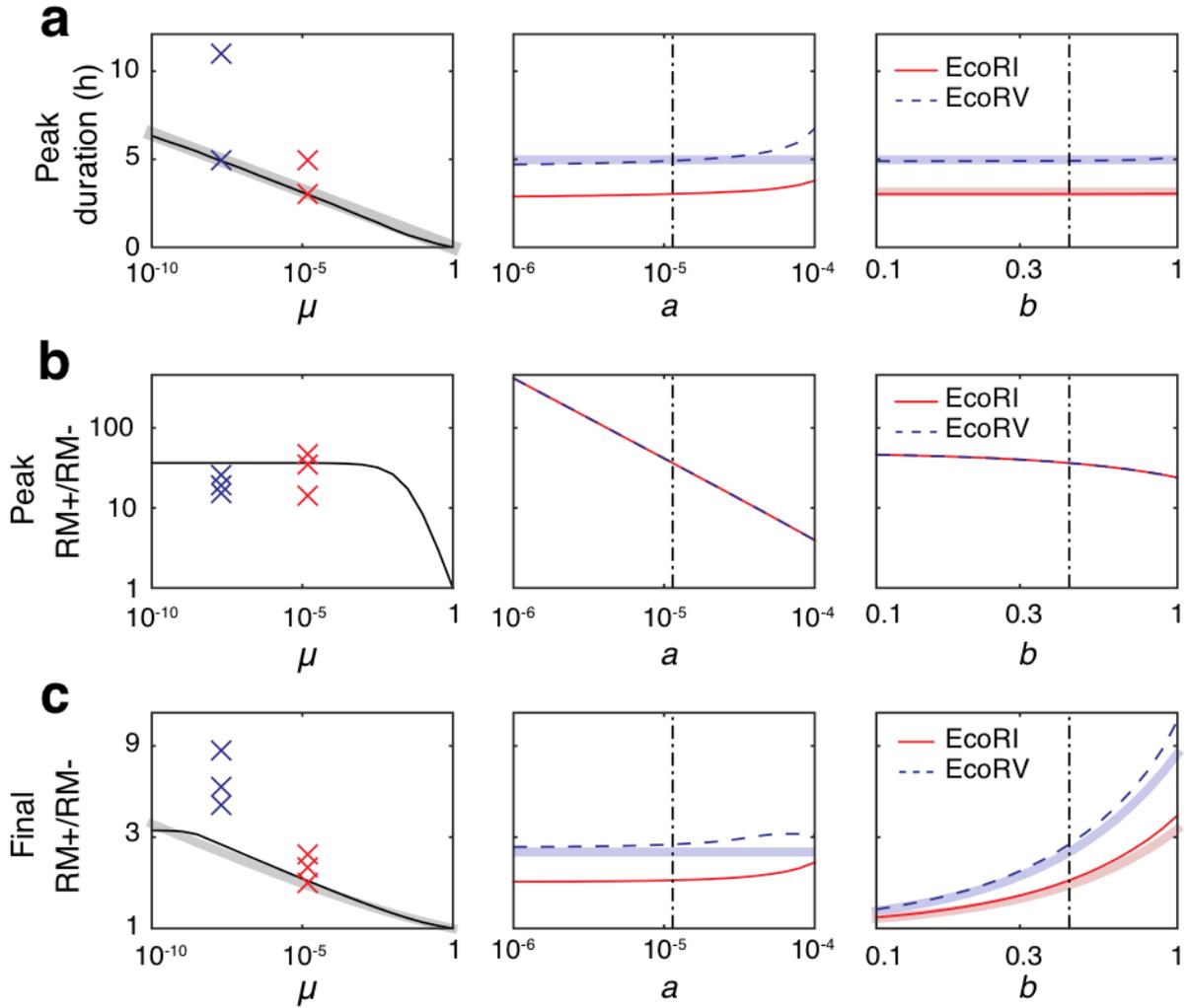
**(B)** Numerical solutions of the deterministic model are shown as bright colored curves. Pale curves represent results of six independent stochastic simulations. Full population dynamics are shown in **Figure 4-10** (deterministic model) and **Figure 4-13** (stochastic model), respectively. The inlay shows the RM+/RM- fold increase as a function of phage escape probability  $\mu$  for different values the parameter  $b$  quantifying the rate with which the probability of lysogeny increases with bacterial density. Shaded areas represent corresponding analytic approximations. Experimentally determined values for EcoRI and EcoRV as shown in **Figure 4-2B** are shown as red and blue points, respectively.



**Figure 4-10: Full numerical solutions of the deterministic model**

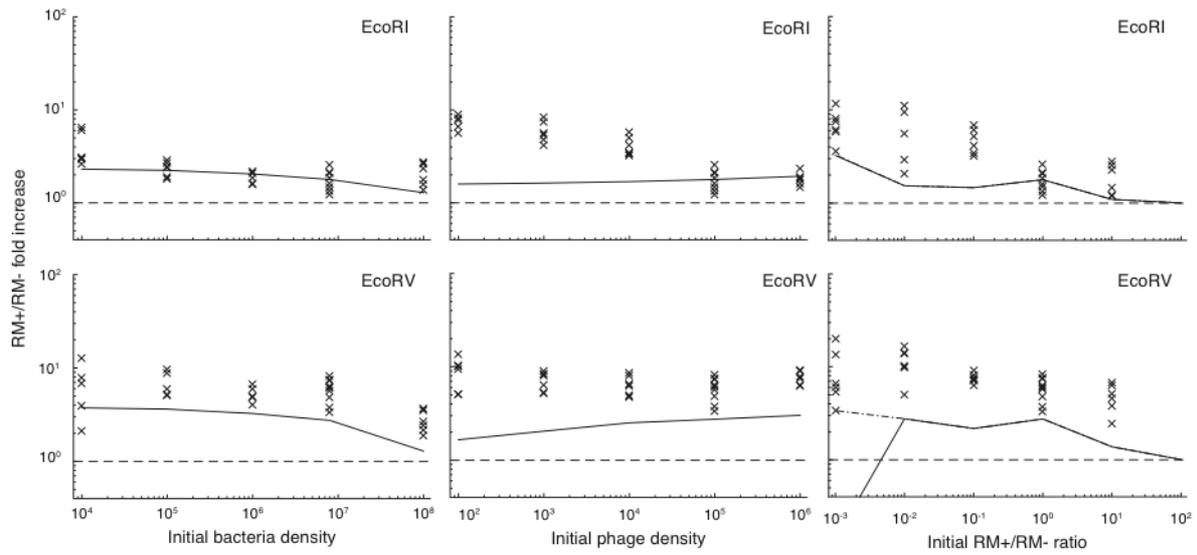
(A) and (B) represent numerical solutions of the model presented in **Figure 4-9B** assuming constant  $\alpha$ .

(C) and (D) represent numerical solutions of the model presented in **Figure 4-9B** with density-dependent  $\alpha$ . The models assume the escape probability ( $\mu$ ) corresponding to EcoRI (A) and (C) or EcoRV (B) and (D).



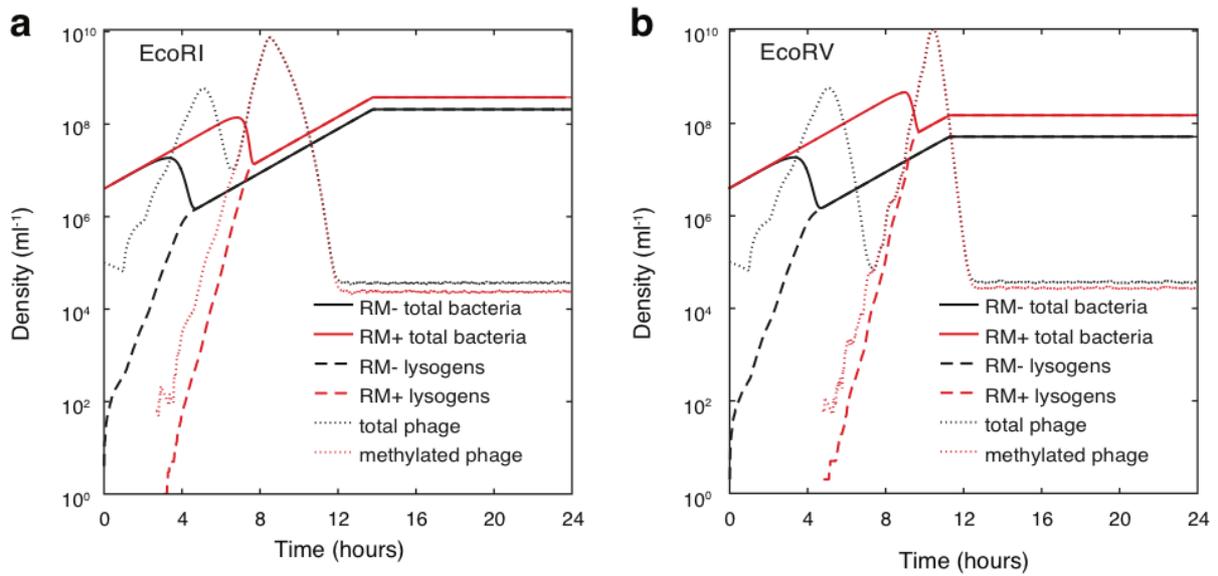
**Figure 4-11: Parameter sensitivity analysis for the mathematical model**

The phage escape probability  $\mu$ , as well as the parameters  $a$  and  $b$  parametrizing the probability of lysogeny as a function of the total biomass were varied in the indicated ranges and their influence on peak duration (a), maximum RM+/RM- ratio (b) as well as the final RM+/RM- ratio (c) was determined as described in Material and Methods. All other parameters were kept at their respective reference values (**Table 4-2**), except for the value of  $\mu$  in the second and third column, which was set to the probability of escape corresponding to EcoRI or EcoRV. Curves represent results of numerical simulations, while the shaded areas (only a and c) represent the analytical approximations. Note that we used a slightly different formula  $\alpha(N) = a N_0^{b_0-b} N^b$  for the probability of lysogeny than the one stated in **Figure 4-9** to allow  $b$  to be varied over a broader range and to simplify comparison of the results. In this formula,  $N_0 = 2.7 \cdot 10^7 \text{ cells/ml}$  and  $b_0 = 0.44$  were used as “reference values” for  $N$  and  $b$  so that the probability of lysogeny at  $t_1$  is approximately equal for all values of  $b$ . Black dashed lines represent experimentally measured values of  $a$  and  $b$ . Red and blue crosses represent experimentally measured values for EcoRI and EcoRV, respectively (**Figure 4-6B**).



**Figure 4-12: Numerical solutions for different initial conditions**

Experimental measurements from **Figure 4-5** are shown as crosses. Solid curves represent solutions of the mathematical model at  $t=24$  hours. Dashed lines correspond to the null hypothesis of equal prophage acquisition of RM+ and RM- bacteria. The dash-dotted line in the lower right plot corresponds to the fold increase at  $t=36$  hours, since in this simulation, unlike in the other cases, equilibrium was not reached in 24 hours.

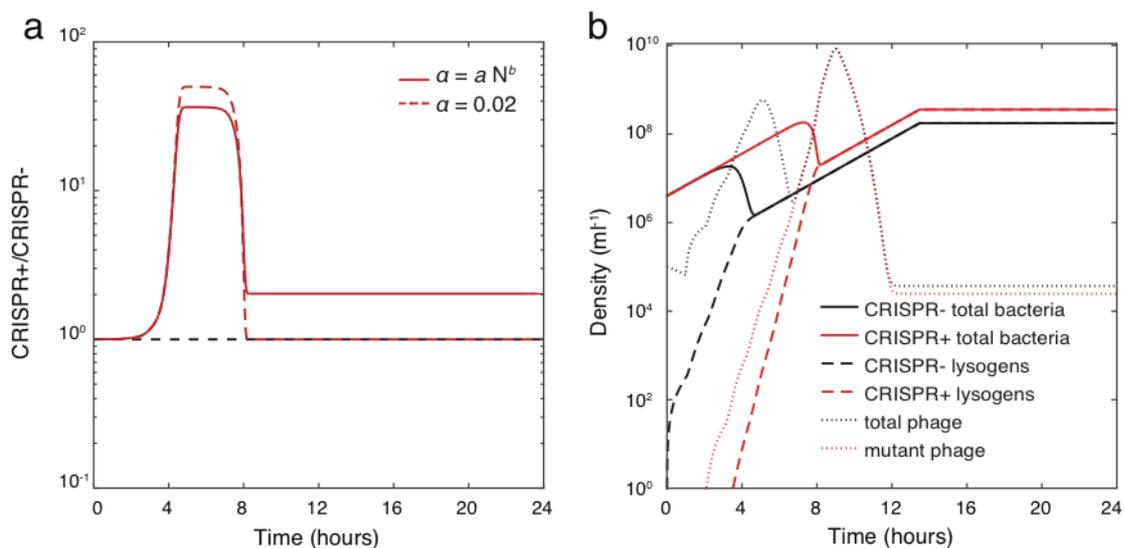


**Figure 4-13: Effect of stochasticity on model dynamics**

Full model dynamics of representative realizations of the stochastic model (**Figure 4-9**) are shown for EcoRI (A) and EcoRV (B). Stochasticity had the strongest direct effect on the time when the first methylated phages appeared. Both for EcoRI and EcoRV, in five out of six stochastic realizations, the final RM+ to RM- ratio was higher than in the deterministic simulation, indicating that stochastic effects tend to slightly increase the final RM+/RM- ratio. In all simulations, we assumed a volume of 1 ml, as compared to 10 ml used in the experiments in order to reduce the computational time.

### 4.3.5 Population dynamics of temperate phages and CRISPR/Cas

We postulate that, as long as the immunity is not absolute and phage escape mutants can be generated (Deveau et al., 2008; Houte, Ekroth, et al., 2016), the population-level advantage for acquiring prophages described here on the example of RM systems would also hold for CRISPR/Cas adaptive immune mechanisms unable to distinguish lytic from lysogenic infections (Edgar & Qimron, 2010) (**Figure 4-14**). Such dynamics could help explain the frequent genomic co-occurrence of prophages with CRISPR/Cas (Touchon, Bernheim, & Rocha, 2016) genes even though, similarly to RM systems, most CRISPR/Cas systems prevent lysogeny at the level of individuals.



**Figure 4-14: Model of interactions between CRISPR/Cas and temperate phage**

**(A)** Numerical solutions of the mathematical model capturing the dynamics of a mixed (CRISPR+ vs. CRISPR-) culture infected with a temperate phage targeted by the CRISPR/Cas. The dashed curve represents numerical solution assuming constant  $\alpha$ . The solid curve represents numerical solution assuming  $\alpha$  to be a function of population density.

**(B)** Full population dynamics of the model depicted in (a) with density-dependent  $\alpha$ . The probability of phage escaping CRISPR (Levin, Moineau, Bushman, & Barrangou, 2013) by mutation is  $\eta = 10^{-6}$ .

#### **4.4 Discussion**

Phages are the most abundant biological entity on Earth (Clokic, Millard, Letarov, & Heaphy, 2011) and, although typically seen as a threat to bacteria, they do more than simply kill their hosts. Phages play a major role as vectors of horizontal gene transfer (Canchaya, Fournous, Chibani-Chennoufi, Dillmann, & Brüssow, 2003), which is a key source of variation for prokaryotic evolution (Ochman, Lawrence, & Groisman, 2000). This dual role is especially important in the case of temperate phages, where each infection can result in the host either dying, or acquiring potentially fitness-augmenting genes. As a key life-history trait, the probability of lysogeny is not constant, but instead depends on a variety of factors such as host physiology (Lieb, 1953; St-Pierre & Endy, 2008) and multiplicity of infection (Kourilsky, 1973; Zeng et al., 2010). For example, bacterial cell size, which decreases with population density (Akerlund, Nordström, & Bernander, 1995), is an important determinant of phage decision making, with small cells preferentially entering lysogeny upon infection (St-Pierre & Endy, 2008). The results presented here demonstrate how, as a result of increasing probability of lysogeny, RM systems as mechanisms of bacterial innate immunity can benefit their hosts without posing a barrier to prophage acquisition. Switching of phage communities from dominance of lysis to lysogeny at high microbial densities has been proposed to occur in a wide range of ecosystems (Knowles et al., 2016), and the general relationship between bacterial density and lysogeny is a matter of current debate (Knowles et al., 2017). Moreover, some temperate phages have been shown to use chemical signaling in order to collectively increase the probability of lysogeny as the number of infections in a population increases (Erez et al., 2017). This being said, the ecological and evolutionary forces determining the probability of lysogeny of temperate phages in general remain poorly understood (Gandon, 2016) and represent an exciting direction for future research.

In addition to acting as a mechanism of host defense, a variety of biological roles have been proposed for RM systems (Vasu & Nagaraja, 2013), including the control of genetic flux (Murray, 2002; Oliveira et al., 2016) and selfish behavior (Naito et al., 1995). Tipping the balance from lysis towards lysogeny in the presence of temperate phages demonstrated here could be another of the multitude of effects these simple genetic elements exert on their hosts. Bioinformatics studies revealed a lack of a negative relationship between the

number of RM systems and presence of prophages in bacteria with large genomes, whereas bacteria with small genomes encoding more RM systems were more likely to carry prophages (Oliveira et al., 2014). Our results offer an explanation of these observations in face of the fact that RM systems do not tolerate lysogenic infections. In addition, temperate phages were shown to avoid restriction sites to a lesser extent than virulent phages (Rocha et al., 2001), suggesting that the two classes of phages are indeed subject to different selection pressures. While numerous experimental studies have focused on interactions between bacterial immune systems and virulent phages (Korona & Levin, 1993; Lenski & Levin, 1985; Levin et al., 2013; Westra et al., 2015), the nature of interactions between bacterial immunity and temperate phages are yet to be explored. The results of this study underscore the critical role of population level dynamics in understanding the evolution and evolutionary consequences of even such simple and mechanistically well-understood genetic elements such as RM systems and phage  $\lambda$ .

## 4.5 *Material and Methods*

### 4.5.1 **RM model assuming constant probability of lysogeny**

The following mathematical model describes population-level dynamics of interactions between temperate phages (Stewart & Levin, 1984) and bacteria with RM systems (Korona & Levin, 1993). We assume a well-mixed habitat of unitary volume, in which RM- ( $i = 0$ ) and RM+ bacteria ( $i = 1$ ) are present as non-lysogens and lysogens at densities  $B_i^-(t)$  and  $B_i^P(t)$  (*cells/ml*), respectively. The phage is present either in a methylated or unmethylated form at density  $P^-(t)$  and  $P^+(t)$  (*phages/ml*), respectively. We assume that, in the absence of phages, all bacteria grow exponentially with the same rate “constant”  $\psi(R(t)) = v_{max} R(t)/(\kappa + R(t))$  ( $h^{-1}$ ), where  $R(t)$  ( $\mu g/ml$ ) is the concentration of the unique limiting resource,  $v_{max}$  ( $h^{-1}$ ) is the maximum growth rate constant and  $\kappa$  ( $\mu g/ml$ ) is the resource concentration at which growth is half-maximal. While a small fitness cost of EcoRI in the medium used has been observed (Pleška et al., 2016) (**Figure 4-2**), this cost is negligible at the time scale of the 24 hour experiments explored by our model. We therefore assume equal values for  $\kappa$  and  $v_{max}$  for all bacteria (**Table 4-2**). The resource is consumed by the bacteria at a rate proportional to their growth rate with a constant of proportionality  $\varepsilon$  ( $\mu g/cell$ ; conversion parameter). Phages adsorb to all bacteria at a rate jointly proportional to their densities with a constant of proportionality  $\delta$  ( $ml/h$ ; adsorption rate constant). For RM- bacteria infected by methylated or unmethylated phages and RM+ bacteria infected by methylated phages, we assume that a fraction  $\alpha$  (*dimensionless*; probability of lysogeny) of infections result in lysogeny, whereas the remaining  $(1 - \alpha)$  of infections lead to lysis and release of  $\beta$  (*phage particles*; burst size) phage particles with the methylation state of the host in which they were produced. Here, we assume that the fraction  $\alpha$  is constant. However, in the following section we drop this assumption and  $\alpha$  will become a function of the total produced biomass. For the sake of simplicity, the model only assumes lysis and lysogeny as possible outcomes of successful infections and does not consider pseudolysogeny, which can occur under conditions of severe nutrient limitation (Ripp & Miller, 1997). For RM+ bacteria infected by unmethylated phages, the fraction  $\mu$  (*dimensionless*; phage escape probability) of infections leads to phage escape, whereas the remaining  $(1 - \mu)$  infections result in phage restriction. We assume equal  $\mu$  for lytic and

lysogenic infections (**Figure 4-1B**). The bacteria in which the infected phage was restricted are assumed to maintain viability. For both RM- and RM+ lysogens, secondary infections do not lead to lysis, and the infecting phages are lost. We assume a time delay of  $\tau$  ( $h$ ; latent period) between adsorption and lysis. Lysogens lyse spontaneously at rate  $\xi$  ( $h^{-1}$ ; induction rate) and produce  $\beta$  particles with the methylation state of the host in which they were produced.

Given these assumptions, the model describing the infection of a mixed population of RM- and RM+ bacteria by initially unmodified phages is given by the following set of delay-differential equations (DDEs):

$$\begin{aligned}
\frac{dR}{dt}(t) &= \underbrace{-\varepsilon \psi(R(t))(B_0^-(t) + B_0^P(t) + B_1^-(t) + B_1^P(t))}_{\text{Consumption of the resource by bacteria}} \\
\frac{dB_0^-}{dt}(t) &= \underbrace{\psi(R(t)) B_0^-(t)}_{\text{Bacterial growth}} - \underbrace{\delta B_0^-(t) (P^-(t) + P^+(t))}_{\text{Lysis and lysogenization}} \\
\frac{dB_0^P}{dt}(t) &= \underbrace{\psi(R(t)) B_0^P(t)}_{\text{Lysogen growth}} + \underbrace{\delta \alpha B_0^-(t) (P^-(t) + P^+(t))}_{\text{Lysogenization}} - \underbrace{\xi B_0^P(t)}_{\text{Induction}} \\
\frac{dB_1^-}{dt}(t) &= \underbrace{\psi(R(t)) B_1^-(t)}_{\text{Bacterial growth}} - \underbrace{\delta B_1^-(t) (\mu P^-(t) + P^+(t))}_{\text{Lysis and lysogenization}} \\
\frac{dB_1^P}{dt}(t) &= \underbrace{\psi(R(t)) B_1^P(t)}_{\text{Lysogen growth}} + \underbrace{\delta \alpha B_1^-(t) (\mu P^-(t) + P^+(t))}_{\text{Lysogenization}} - \underbrace{\xi B_1^P(t)}_{\text{Induction}} \\
\frac{dP^-}{dt}(t) &= \underbrace{\delta (1 - \alpha) \beta B_0^-(t - \tau) (P^-(t - \tau) + P^+(t - \tau))}_{\text{Lysis}} + \underbrace{\xi \beta B_0^P(t)}_{\text{Induction}} \\
&\quad - \underbrace{\delta P^-(t) (B_0^-(t) + B_0^P(t) + B_1^-(t) + B_1^P(t))}_{\text{Adsorption}} \\
\frac{dP^+}{dt}(t) &= \underbrace{\delta (1 - \alpha) \beta B_1^-(t - \tau) (\mu P^-(t - \tau) + P^+(t - \tau))}_{\text{Lysis}} + \underbrace{\xi \beta B_1^P(t)}_{\text{Induction}} \\
&\quad - \underbrace{\delta P^+(t) (B_0^-(t) + B_0^P(t) + B_1^-(t) + B_1^P(t))}_{\text{Adsorption}}
\end{aligned}$$

Note, that the rate of phage production by lysis at any given time  $t$  depends on the density of non-lysogens and phages at time  $t - \tau$  due to the latent period between adsorption and lysis. In all simulations, we assume that phages are added to the system at  $t = 0$ , i.e. we assume  $P^-(t < 0) = P^+(t < 0) = 0$ . The parameters of the model (**Table 4-2**) were

estimated experimentally as described below. Initial conditions are given in **Error! eference source not found.** Deterministic simulations as well as the numeric parameter sensitivity analysis were performed using Matlab R2015a (The MathWorks, Inc., Natick, Massachusetts, United States) using the dde23 solver.

**Table 4-2: Model parameters**

Parameter	Symbol	Estimated value	N replicas
Maximum growth rate <sup>a</sup>	$v_{max}$	$0.55 h^{-1}$	8 (average of 10 each)
Monod constant <sup>b</sup>	$\kappa$	$1 \mu g/ml$	
Conversion parameter	$\varepsilon$	$5 \cdot 10^{-7} \mu g/bacterium$	5
Adsorption rate	$\delta$	$4.1 \cdot 10^{-8} ml/h$	3
Latent period	$\tau$	$1 h$	3
Burst size	$\beta$	$63 particles$	3
Phage escape probability			
EcoRI	$\mu_{EcoRI}$	$1.37 \cdot 10^{-5}$	7
EcoRV	$\mu_{EcoRV}$	$1.96 \cdot 10^{-8}$	7
Probability of lysogeny <sup>6</sup>	$\alpha$	0.02	4
Induction rate	$\xi$	$2.4 \cdot 10^{-5} h^{-1}$	3
Parameters for biomass	$a$	$1.14 \cdot 10^{-5}$	
dependent prob. of lysogeny	$b$	0.44	
Mutation probability	$\eta$	$10^{-6}$	(Levin et al., 2013)

<sup>a</sup> Growth rates of RM- and RM+ (EcoRI/EcoRV) bacteria, as well as  $\lambda kan$  lysogens did not differ significantly when tested by linear regression, with growth rate as a continuous dependent variable and strain identity as an independent categorical variable, comparing individual strains to the control (RM-) P values were 0.81, 0.67 and 0.64 for EcoRI, EcoRV and  $\lambda kan$  lysogens, respectively.

<sup>b</sup> Arbitrary value. Different values do not affect the result qualitatively.

**Table 4-3: Model initial conditions**

State	Symbol	Initial condition
RM- bacteria, non-lysogens	$B_0^-$	$4 \cdot 10^6$ cells/ml
RM- bacteria, lysogens	$B_0^P, B_0^M$	0 cells/ml
RM+ bacteria, non-lysogens	$B_1^-$	$4 \cdot 10^6$ cells/ml
RM+ bacteria, lysogens	$B_1^P$	0 cells/ml
CRISPR+ bacteria, non-lysogens	$B_2^-$	$4 \cdot 10^6$ cells/ml
CRISPR+ bacteria, lysogens	$B_2^M$	0 cells/ml
Phages, unmethylated	$P^-$	$10^5$ phages/ml
Phages, methylated	$P^+$	0 phages/ml
Phages, mutated	$P^M$	0 phages/ml
Resources for growth	$R$	400 $\mu\text{g/ml}$

#### 4.5.2 RM model assuming non-constant probability of lysogeny

The experiments presented in **Figure 4-9A** showed a clear correlation between the probability of lysogeny  $\alpha$ , and the bacterial density at the time of infection. Since the underlying mechanism responsible for this correlation is unknown, we incorporate this dependency into the mathematical model by replacing the constant probability of lysogeny by a function  $\alpha(N(t)) = a N^b(t)$  of the total produced biomass  $N(t)$ , which is given by the solution of the additional differential equation

$$\frac{dN}{dt}(t) = \psi(R(t)) (B_0^-(t) + B_0^P(t) + B_1^-(t) + B_1^P(t)),$$

with initial condition  $N(0) = B_0^-(0) + B_0^P(0) + B_1^-(0) + B_1^P(0)$ . Unlike the total cell density,  $N(t)$  is monotonically increasing in time, and is less sensitive to specific modeling decisions, such as the time point until which infected bacteria influence the probability of lysogeny of other infection events. Furthermore, because  $R(t) + \varepsilon N(t)$  is always constant, rendering the probability of lysogeny a function of the total produced biomass is equivalent to rendering it a function of the limiting resource.

### 4.5.3 Multiplicity-of-infection-related effects

Besides physiological state of infected bacteria, the average number of phages infecting a single bacterium, also known as multiplicity of infection (MOI), affects the probability of lysogeny (Kourilsky, 1973). However, to our knowledge, there is no established method of estimating the MOI in population dynamics experiments. Note that in dynamical experiments, MOI is not equal or well approximated by the phage/bacteria ratio, as is the case in short-term experiments, in which all phages adsorb to bacteria synchronously and neither phages nor bacteria multiply during the experiment. In contrast to such setting, phages in our experiments do not only adsorb, but are also continuously produced by previously infected bacteria and the total number of infections is thus not bounded by the total number of free phages at any given time. Furthermore, because phages adsorb to bacteria continuously, it is likely that after the first phage adsorbs and infects a bacterium, subsequent infections of the same bacterium can affect the outcome only if they occur before the infection outcome was decided by the first adsorbing phage. However, it is not known how long this time interval is, neither how do infections at different times affect the outcome. Due to these conceptual challenges regarding the definition of the MOI in population dynamics experiments, our model does not include MOI as a parameter affecting the probability of lysogeny. Understanding how the MOI effects combine with the effect of host physiology to influence the population dynamics of lysogeny remains a goal for future research.

### 4.5.4 CRISPR model

The presented model is a modified version of the model capturing interactions between CRISPR-carrying bacteria and virulent phage (Levin et al., 2013). In this model, phages can escape the effects CRISPR by genetic mutations in the DNA sequence targeted by the CRISPR system, which stands in contrast to epigenetic escape by methylation considered previously. We assume that the phage is present either as the wild type or as a CRISPR escape mutant at densities  $P^-(t)$  and  $P^M(t)$  (phages/ml), respectively. CRISPR+ ( $i = 2$ ) and CRISPR- bacteria ( $i = 0$ ) are present as non-lysogens or lysogens. Changes in density of non-lysogens  $B_i^-$  are modeled as described in the RM model, whereas we distinguish between

lysogens  $B_i^P$  carrying the wild-type prophage, and lysogens  $B_i^M$  (*cells/ml*) carrying the escaped mutant prophage. We assume that infections of all CRISPR+ bacteria with wild-type phages lead to cleavage of phage DNA, i.e. the infecting phages are lost. All other infections are assumed to lead to lysis or lysogeny as described above. When bacteria lyse, the majority of phages produced are either wild type if the infecting phage was wild type, or CRISPR escape mutants otherwise. However, we assume that a fraction  $\eta$  (*dimensionless*; mutation rate) of the produced phages escape by mutation and become insensitive to CRISPR. With all other interaction dynamics assumed to be the same as in the RM model, we obtain the following set of equations:

$$\begin{aligned} \frac{dR}{dt}(t) &= \underbrace{-\varepsilon \psi(R(t))(B_0^-(t) + B_0^P(t) + B_0^M(t) + B_2^-(t) + B_2^M(t))}_{\text{Consumption of the resource by bacteria}} \\ \frac{dB_0^-}{dt}(t) &= \underbrace{\psi(R(t)) B_0^-(t)}_{\text{Bacterial growth}} - \underbrace{\delta B_0^-(t) (P^-(t) + P^M(t))}_{\text{Lysis and lysogenization}} \\ \frac{dB_0^P}{dt}(t) &= \underbrace{\psi(R(t)) B_0^P(t)}_{\text{Lysogen growth}} + \underbrace{\delta \alpha B_0^-(t) P^-(t)}_{\text{Lysogenization}} - \underbrace{\xi B_0^P(t)}_{\text{Induction}} \\ \frac{dB_0^M}{dt}(t) &= \underbrace{\psi(R(t)) B_0^M(t)}_{\text{Lysogen growth}} + \underbrace{\delta \alpha B_0^-(t) P^M(t)}_{\text{Lysogenization}} - \underbrace{\xi B_0^M(t)}_{\text{Induction}} \\ \frac{dB_2^-}{dt}(t) &= \underbrace{\psi(R(t)) B_2^-(t)}_{\text{Bacterial growth}} - \underbrace{\delta B_2^-(t) P^M(t)}_{\text{Lysis and lysogenization}} \\ \frac{dB_2^M}{dt}(t) &= \underbrace{\psi(R(t)) B_2^M(t)}_{\text{Lysogen growth}} + \underbrace{\delta \alpha B_2^-(t) P^M(t)}_{\text{Lysogenization}} - \underbrace{\xi B_2^M(t)}_{\text{Induction}} \\ \frac{dP^-}{dt}(t) &= \underbrace{\delta (1 - \alpha) \beta (1 - \eta) B_0^-(t - \tau) P^-(t - \tau)}_{\text{Lysis}} + \underbrace{\xi \beta (1 - \eta) B_0^P(t)}_{\text{Induction}} \\ &\quad - \underbrace{\delta P^-(t) (B_0^-(t) + B_0^P(t) + B_0^M(t) + B_2^-(t) + B_2^M(t))}_{\text{Adsorption}} \\ \frac{dP^M}{dt}(t) &= \underbrace{\delta (1 - \alpha) \beta (B_0^-(t - \tau) + B_2^-(t - \tau)) P^M(t - \tau)}_{\text{Lysis}} + \underbrace{\xi \beta (B_0^M(t) + B_2^M(t))}_{\text{Induction}} \\ &\quad - \underbrace{\delta P^M(t) (B_0^-(t) + B_0^P(t) + B_0^M(t) + B_2^-(t) + B_2^M(t))}_{\text{Adsorption}} \\ &\quad + \underbrace{\delta (1 - \alpha) \beta \eta B_0^-(t - \tau) P^-(t - \tau) + \xi \beta \eta B_0^P(t)}_{\text{Modification}}. \end{aligned}$$

Note that, since we assume that all wild-type phages adsorbed by CRISPR+ bacteria are

cleaved, the density of CRISPR+ lysogens carrying the wild-type phage is always zero ( $B_2^P(t) = 0$ ). The case when the probability of lysogeny is not constant but depends on the total produced biomass is treated in the same way as described in the RM model.

#### 4.5.5 Parameter sensitivity analysis

In order to understand how the individual key properties of RM systems and temperate phages affect the population dynamics, we analyzed how these depend on the phage escape probability  $\mu$ , and on the parameters  $a$  and  $b$  parametrizing the probability of lysogeny  $\alpha(N(t))$  as a function of the total produced biomass  $N(t)$ . To allow  $b$  to be varied over a broader range and to simplify comparison, we used a slightly different formula

$$\alpha(N) = a N_0^{b_0-b} N^b = a N_0^{b_0} \left(\frac{N}{N_0}\right)^b$$

**Figure 4-9.** The “reference values”  $N_0 = 2.7 \cdot 10^7 \text{ cells/ml}$  and  $b_0 = 0.44$  were chosen such that the probability of lysogeny at the time of the first wave of infection is approximately the same for all values of  $b$ . Note, that the two formulas are identical when  $b = b_0$ .

We characterized the model dynamics using three key properties (**Figure 4-11**, compare with **Figure 4-6**): the time  $t_1$  quantifies the time point when a sufficient number of unmethylated phages have accumulated such that the majority of the RM- bacteria either lyse or become lysogens. At this point, the ratio between RM+ and RM- bacteria, which is initially approximately one, quickly increases and approaches a maximum value, which we refer to as the peak RM+/RM- ratio  $r_p$ . At time  $t_2$ , the density of methylated phages becomes sufficiently high such that the majority of RM+ cells either lyses or becomes lysogenic. We refer to the time difference  $\Delta t = t_2 - t_1$  as the peak duration. After  $t_2$ , the ratio between RM+ and RM- bacteria drops and approaches a final value, to which we refer as the final RM+/RM- ratio  $r_f$ . For each combination of  $\mu$ ,  $a$  and  $b$ , we run a separate numerical simulation and identified  $t_i$  as the time point at which

$$\frac{B_i^-(t)}{B_i^-(t) + B_i^P(t)} = \frac{1}{2} r_p$$

as the maximum value of  $r(t) = \frac{B_1^-(t) + B_1^P(t)}{B_0^-(t) + B_0^P(t)}$  between  $t_1$  and  $t_2$ , and  $r_f$  as the value of  $r(t)$  for

$t \gg t_2$ . In order to compare these characteristic properties of the simulated model dynamics with the (comparatively noisy) experimentally observed dynamics, we fitted piecewise constant functions

$$\hat{r}(t) = \begin{cases} 1 & \text{for } t < t_1 \\ \hat{r}_p & \text{for } t_1 \leq t < t_2 \\ \hat{r}_f & \text{for } t \geq t_2 \end{cases}$$

to the experimental data (**Figure 4-6B**). Since all values of  $t_1$  or  $t_2$  lying between the same adjacent experimental sampling times result in the same fitting errors, we restricted  $t_1$  and  $t_2$  to values exactly in the middle of adjacent sampling times.

Despite the relatively large number of species in the model, it is possible to derive a surprisingly good analytic approximation of the peak duration and the final RM+/RM- ratio. If the initial RM- and RM+ bacterial densities are similar and sufficiently high, the growth rates of both the unmethylated and the methylated phage are substantially higher than the bacterial growth rate. As a consequence, most RM- and RM+ bacteria get infected during rather short time intervals around  $t_1$  and  $t_2$ , respectively. Thus, it is sufficient to consider the probabilities of lysogeny  $\alpha_1 = \alpha(N(t_1))$  and  $\alpha_2 = \alpha(N(t_2))$  at  $t_1$  and  $t_2$ , and the final RM+/RM- ratio can be approximated by  $r_f \approx \frac{\alpha_2 B_1^-(0)}{\alpha_1 B_0^-(0)}$ . Furthermore, until shortly before  $t_2$ , the dynamics of RM- lysogens, RM- non-lysogens, and RM+ non-lysogens, as well as of the unmethylated phage are approximately independent of the dynamics of the methylated phage. Before  $t_1$ , the density of methylated phage is substantially below that of the unmethylated phage. After  $t_1$ , there are only few RM- non-lysogens present and production of unmethylated phage is thus negligible. If we consider  $P^-(t)$ ,  $B_0^-(t)$ ,  $B_0^P(t)$  and  $B_1^-(t)$  as fixed (time-dependent) functions, rather than states in the DDE for the methylated phage, and if we neglect the (relatively small) influence of RM+ lysogens on the methylated phage dynamics before  $t_2$ , the DDE of the methylated phage becomes linear. Thus, the methylated phage dynamics are approximately proportional to the phage escape probability until shortly before  $t_2$ . As a consequence, when we scale the phage escape probability by a factor  $f < 1$ , the peak duration grows by the time required for the methylated phage to grow by a factor of  $1/f$ . Since, for  $\mu = 1$ , the peak duration is zero, we obtain

$$\Delta t \approx \frac{1}{k_{max}^+} \log \frac{1}{\mu},$$

with  $k_{max}^+ = \frac{1}{\tau} \log(\beta) - v_{max}$  being the maximal growth rate of the methylated phage after  $t_1$  and before  $t_2$ , which is approached at sufficiently high initial densities of RM+ bacteria.

Given the peak duration, we can approximate the increase in the total biomass between  $t_1$  and  $t_2$ , and obtain:

$$N(t_2) \approx \left(1 + \frac{B_1^-(0)}{B_0^-(0) + B_1^-(0)} (e^{v_{max} \Delta t} - 1)\right) N(t_1).$$

This allows us to calculate the ratio between the probabilities of lysogeny  $\alpha_1$  and  $\alpha_2$ . The final ratio between RM+ and RM- bacteria can then be approximated as:

$$r_f \approx \left(1 + \frac{B_1^-(0)}{B_0^-(0) + B_1^-(0)} \left(\mu^{-\frac{v_{max}}{k_{max}^+}} - 1\right)\right)^b \frac{B_1^-(0)}{B_0^-(0)}.$$

Even though these formulas for the peak duration and the final RM+/RM- ratio are based on many simplifications, they are in surprisingly good agreement with the numeric results (**Figure 4-9B** and **Figure 4-11**). They indicate that the peak duration is approximately proportional to the logarithm of the inverse of the phage escape probability  $\mu$ , but depends only little on the probability of lysogeny  $\alpha$ , given that the latter does not become too high. On the other hand, the final RM+/RM- ratio grows approximately exponentially with the parameter  $b$ , but depends only little on the parameter  $a$  (recall that the probability of lysogeny is given by  $\alpha(N(t)) = a N^b(t)$ ). Finally, for small enough values of  $\mu$ , the final RM+/RM- ratio is approximately proportional to  $\left(\frac{1}{\mu}\right)^b \frac{v_{max}}{k_{max}^+}$ . The final RM+/RM- ratio saturates for very small phage escape probabilities because the peak duration becomes long enough such that bacteria have already reached stationary phase before  $t_2$ .

#### 4.5.6 Stochastic model

We implemented a stochastic version of the RM model in which the dynamics are described by a series of “events” between individual bacteria and phages taking place with a given probability per unit time, rather than as continuous “reactions” with deterministic rates as in the DDE model. Thus, if the variation in the time when the first methylated phages occurred in the replicas of the competition experiment for EcoRV (**Figure 4-6B**) was a consequence of the low phage escape probability, we would expect to see a similar variation in individual simulations of the stochastic model. For this model, we assumed the same interactions between species and the same parameter set identified for the

deterministic model (**Table 4-2**). While we used a total medium volume of 10 ml in the experiments, we assumed a volume of only 1 ml in these simulations to reduce computational time. Note that the effect of stochastic noise due to small reaction propensities is usually the higher the smaller the volume. Simulations were performed using Gillespie's (Gillespie, 1977) algorithm as implemented in the software Dizzy (Ramsey, Orrell, Bolouri, & Others, 2005). Small modifications to the source code of Dizzy were necessary to allow for the specific type of delayed reactions in the model, to speed up the simulations, and to reduce the memory requirements. Specifically, in the original implementation, every delayed reaction is split into two components: a first component immediately decreasing the absolute numbers of the reactants, and a second component increasing the absolute numbers of the products exactly after the defined delay. As a consequence, for each firing of the reaction, the precise time of the second component has to be kept in memory, which significantly increases the memory footprint and decreases speed. We modified this original algorithm such that second components closer together than 10 s are pooled into a common second component "firing multiple times". Only with these modifications to the original algorithm, and only by assuming a total volume of 1 ml instead of 10 ml, it was possible to run the stochastic simulations in reasonable time (between one and two days per simulation on a 3.3GHz PC with 8GB RAM).

#### **4.5.7 Parameter estimation**

**Maximum growth rate:** Overnight cultures were diluted 1:250 in a flat-bottom 96-well plate into fresh medium (200  $\mu$ l total volume). The plate was continuously shaken inside the Synergy H1 Multi-Mode Reader (Bio-Tek) and OD600 was measured at ten-minute intervals for ten hours. Growth-rates were calculated from the background-subtracted values of OD600 as the time derivative of  $\ln$  OD600) during 90 minutes of exponential growth. The outmost wells were used for background subtraction. **Adsorption rate:** Approximately  $10^6$  pfu/ml of  $\lambda$  kan were added to a growing bacterial culture and the density of non-adsorbed phage was measured in 10-minute intervals for 30 minutes. At each time point, 1 ml of the sample was filtered (0.2  $\mu$ m) and serial dilutions were plated to estimate the density of free phage. Bacterial density was estimated independently. The adsorption rate was calculated as the time derivative of free phage density, normalized by bacterial density. **Latent period**

and **burst size** were estimated as described previously (Ellis & Delbrück, 1939). The **conversion parameter** was calculated as  $e = R/D$ , where  $R$  is the resource density in fresh medium ( $\mu\text{g/ml}$ ) and  $D$  is the bacterial density reached in a fully-grown culture. The **induction rate** was calculated from the equilibrium density of free phage  $\bar{P}$  in fully-grown cultures of  $\lambda$  *kan* lysogens as  $i = \delta / \beta \cdot \bar{P}$ , where  $\delta$  is the adsorption rate and  $\beta$  is the burst size. Adsorption rate, burst size, latent period and the probability of lysogeny were estimated using  $\lambda$  *kan* and exponentially growing *E. coli* MG1655 (four hour incubation after 1:100 dilution).

#### 4.5.8 Media and growth conditions

Unless otherwise stated, bacteria were grown in M9 maltose medium (1x M9 salts (12.8 g/l  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 3 g/l  $\text{KH}_2\text{PO}_4$ , 0.5 g/l NaCl, 1 g/l  $\text{NH}_4\text{Cl}$ ), 0.4% maltose, 2 mM  $\text{MgSO}_4$ , 0.1 mM  $\text{CaCl}_2$ ) at 37 °C with vigorous shaking. M63 medium (1x M63 salts (2 g/l  $(\text{NH}_4)_2\text{SO}_4$ , 13.6 g/l  $\text{KH}_2\text{PO}_4$ , 0.5 mg/l  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ ), 0.4% maltose, 2 mM  $\text{MgSO}_4$ , 0.2 mM  $\text{CaCl}_2$ ) was used for experiments shown in **Figure 4-3**. Adding 0.2 mM  $\text{CaCl}_2$  into the M63 medium was necessary for successful infection by  $\lambda$  *kan*. We used defined media for the competition experiments to increase reproducibility and provide well-defined physiological conditions, in which bacteria compete for a single limiting resource (maltose), which is also an assumption of our mathematical model. In all competition experiments, RM+ and RM- bacteria were  $ara^- cat^+$  and  $ara^+ cat^-$ , respectively. The maintenance of EcoRI and EcoRV plasmids was selected for with 100  $\mu\text{g/ml}$  ampicillin. For the estimation of bacterial density, 10-100  $\mu\text{l}$  of a diluted sample was plated on Petri dishes and spread with sterile glass beads. LB plates (1% Agar) were used for estimating the total bacterial density. Tetrazolium-arabinose (TA) agar plates (1% tryptone, 0.1% yeast extract, 0.5% NaCl, 1% agar, 1% arabinose, 0.005% tetrazolium (Sigma-Aldrich)) were used as indicator plates to estimate density of  $ara^+ cat^-$  and  $ara^- cat^+$  bacteria. When the frequency of one of the two types was below 1:50, M9 minimal arabinose plates (1x M9 salts, 0.4% arabinose, 2 mM  $\text{MgSO}_4$ , 0.1 mM  $\text{CaCl}_2$ , 1% agar) and LB plates supplemented with chloramphenicol (50  $\mu\text{g/ml}$ ) were used to estimate densities of  $ara^+ cat^-$  and  $ara^- cat^+$  bacteria, respectively. Kanamycin (20  $\mu\text{g/ml}$ ) was added to LB, TA and chloramphenicol plates to select for lysogens. Phage plates (1% tryptone, 0.1% yeast extract, 0.8% NaCl, 1% agar, 0.01% glucose, 0.2 mM  $\text{CaCl}_2$ ) were used for estimating phage

density. Lawns of RM- or RM+ bacteria prepared by mixing 100  $\mu$ l of overnight cultures in 3 ml of phage soft agar (1% tryptone, 0.1% yeast extract, 0.8% NaCl, 0.7% agar, 0.01% glucose, 0.2 mM CaCl<sub>2</sub>) were used to estimate densities of total and methylated phage, respectively. Dilutions were performed in 96-well plates with SM buffer. All plates were incubated at 37 °C overnight, with the exception of M9 minimal arabinose plates, which were incubated for at least 36 hours before counting. Colonies and plaques were counted manually. Several dilutions were plated for each measurement and, whenever possible, plates with 20-200 colonies/plaques were counted.

#### **4.5.9 Lysate preparation**

Phage lysates were prepared by plate lysis. Specifically, individual phage plaques were picked with a sterile pipette tip, resuspended in 3 ml of phage soft agar together with 100  $\mu$ l of overnight bacterial culture and plated on top of phage plates. The plates were then incubated at 37°C overnight. The soft agar was scraped with a sterile microscope glass slide, resuspended in 10 ml of SM buffer (100 mM NaCl, 8 mM MgSO<sub>4</sub>, 200 mM Tris-Cl (pH 7.5)) with a few drops of chloroform to kill the residual bacteria. The lysates were then centrifuged to remove the leftover agar, sterilized by filtration (0.2  $\mu$ m) and stored at 4 °C. The  $\lambda$  *kan* lysates used for competition experiments were grown on lawns of MG1655. The same lysates were used to measure the *eop* and *eol* of EcoRI, EcoRV, EcoPI, EcoP15I and EcoVIII RM systems. Because the MG1655 strain carries a type I RM system (EcoKI), the lysates used to measure the *eop* and *eol* of EcoAI, EcoBI, EcoKI and EcoR124I were obtained by growing  $\lambda$  *kan* on *E. coli* C-1, a strain devoid of methyltransferases. These lysates were also used to measure the *eop* and *eol* of EcoRII, a RM system whose recognition sequence overlaps with the solitary methyltransferase *dcm* present in MG1655 (Takahashi, Naito, Handa, & Kobayashi, 2002). The lysates used to measure the *eop* and *eol* of EcoKMcrBC, a type IV RM system cleaving DNA with the PvuII methylation pattern were obtained by growing  $\lambda$  *kan* on the MP084 strain carrying a plasmid with this RM system (pPvuII 3.4 (Blumenthal, Gregory, & Cooperider, 1985)).

#### 4.5.10 Measuring efficiencies of plating and lysogenization

**eop:** serially diluted  $\lambda$  *kan* lysate was mixed with 0.1 ml of an overnight culture (LB) in 3 ml of phage soft agar and spread on phage plates. *eop* was calculated as the relative ratio of the plaque-forming units (*pfu*) obtained on lawns of RM+ and RM- bacteria ( $eop = pfu_{RM+}/pfu_{RM-}$ ). **eol:** 0.5 ml of an overnight culture ( $\sim 10^8$  bacteria) was mixed with the  $\lambda$  *kan* lysate to reach phage/bacteria ratio of 1. Samples were incubated on ice for 30 min to allow for phage adsorption, then incubated for additional 30 min at 37 °C to allow for phage infection and expression of kanamycin resistance. In order to prevent formation of lysogens on agar plates, samples were washed once with SM buffer to wash away the non-adsorbed phage. Serial dilutions were plated on LB plates with 20  $\mu$ g/ml kanamycin to estimate the density of lysogenic colony-forming units (*lcfu*). *eol* was calculated as the relative ratio of *lcfu* obtained for RM+ and RM- bacteria ( $eol = lcfu_{RM+}/lcfu_{RM-}$ ). The following strains and plasmids were used in these experiments: NK354 (EcoAI), WA251(EcoBI), MG1655 (EcoKI and EcoKMcrBC), NK402 (EcoR124I), pBR322 $\Delta$ Ptet EcoRI (R+M+) (EcoRI), pMPR001 (EcoRII), pBR322 $\Delta$ Ptet EcoRV (R+M+) (EcoRV), pRR0 (EcoVIII), pNR201 (EcoP1), pNR301 (EcoP15I).

#### 4.5.11 Competition experiments

All competitions were between MG1655 (*ara*<sup>+</sup> *cat*<sup>-</sup>) and MP085 (*ara*<sup>-</sup> *cat*<sup>+</sup>) strains. These strains carried either the empty pBR322 $\Delta$ Ptet plasmid (RM-), or one of the plasmids carrying an RM system: pBR322 $\Delta$ Ptet EcoRI (R+M+), or pBR322 $\Delta$ Ptet EcoRV (R+M+) (Pleška et al., 2016). Individual colonies from overnight plates were inoculated into 2 ml of medium and grown for 24 hours at 37 °C with vigorous shaking. The two strains were then mixed in a desired ratio and diluted 1:100 into fresh medium. Mixtures were first grown for 24 hours at 37 °C in the absence of phage and then diluted into two separate tubes with 2 ml of fresh medium to reach the desired bacteria density.  $\lambda$  *kan* lysates grown on MG1655 were diluted in SM buffer and 2  $\mu$ l of the corresponding dilution was added to reach the desired phage density. When the ratio between the two types was close to 1, the density of RM+ and RM- lysogens was estimated simultaneously by plating diluted cultures on TA plates supplemented with 20  $\mu$ g/ml Kanamycin. Otherwise, the densities were estimated

individually by plating the samples either on LB plates with both kanamycin and chloramphenicol (RM+ lysogens), or minimal arabinose plates with kanamycin (RM- lysogens). The RM+/RM- fold increase was calculated as  $\left(\frac{RM_{\lambda}^{+}(24)}{RM_{\lambda}^{-}(24)}\right) / \left(\frac{RM^{+}(0)}{RM^{-}(0)}\right)$ , where  $\frac{RM_{\lambda}^{+}(24)}{RM_{\lambda}^{-}(24)}$  is the ratio of lysogens obtained 24 hours after infection and  $\left(\frac{RM^{+}(0)}{RM^{-}(0)}\right)$  is the ratio of sensitive bacteria at the beginning of the experiment.

In experiments presented in **Figure 4-4**, the cultures were serially transferred every 24 hours by 1:100 dilution into 2 ml of medium and bacterial densities were estimated by plating on TA plates. Phage density was estimated by spotting 5  $\mu$ l of serially diluted sample on lawns of MG1655. After the first and last day of the experiment, five colonies of both types from each culture were tested for their ability to grow on kanamycin to test for lysogeny. For the experiments depicted in **Figure 4-6**, mixed cultures were diluted 1:100 into 10 ml of fresh medium and infected with  $\lambda$  *kan* to reach the density of approximately  $10^5$  pfu/ml. Cultures were incubated inside an Innova® 3100 (New Brunswick™) water bath at 37 °C with constant shaking. Samples (200  $\mu$ l) were taken at one-hour intervals.

#### 4.5.12 Measuring the probability of lysogeny

Overnight cultures were diluted 1:100 and grown for 4 hours at 37 °C with vigorous shaking. 1 ml samples were infected with  $\lambda$  *kan* at phage/bacteria ratio equal to 0.1. Samples were incubated on ice for 30 min to allow for phage adsorption, then incubated for additional 30 min at 37 °C to allow for phage infection and expression of kanamycin resistance. Infected bacteria were washed with SM buffer to remove the non-adsorbed phage and serial dilutions were plated on a) LB kanamycin plates to estimate the density of *lcfu*, and b) lawns of sensitive bacteria to estimate the density of infective centers. To make sure that majority of the infections are lytic, the plates to estimate the density of infected centers were exposed to UV before incubation. The probability of lysogeny was calculated as  $\alpha = lcfu/pfu$ . In experiments depicted in **Figure 4-9B**, overnight cultures were diluted 1:100 into fresh medium and incubated at 37 °C with vigorous shaking. Samples were taken at two-hour intervals and  $\alpha$  was measured as above. At each time-point,  $\lambda$  *kan* was added to reach phage/bacteria ratio of 0.1.

#### 4.5.13 Statistical analysis

All statistical tests were performed using Matlab R2015a (The MathWorks, Inc., Natick, Massachusetts, United States). In **Figure 4-1B**, the eop and eol were compared for each RM system individually using Welch's t-test ( $H_0: eol \leq eop$ ). The level of significance was adjusted using Bonferroni correction for multiple comparisons:  $\alpha^* = \frac{0.05}{10} = 0.005$ . Linear regression models were fit using the `fitlm()` command. RM+/RM- fold increase data were log-transformed before analysis. Normal distribution of errors was verified by residual analysis. The symbols depicting levels of significance correspond to: ns (not significant)  $p > 0.05$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

#### 4.5.14 Strain and plasmid construction

**MP084** was constructed by P1 transduction of the  $\Delta mcrB::kanR$  allele from JW5871 (KEIO collection (Baba et al., 2006)) into MG1655, followed by removal of the kanamycin resistance marker using the pCP20 plasmid (Haldimann & Wanner, 2001). **TB357** was constructed by recombineering of the  $\Delta araA::kanR-frt$  PCR fragment (amplified by *5\_araA\_KO* and *3\_araA\_KO* from pKD13) into MG1655. The KanR marker was then flipped out using pCP20. **MP085** was constructed by integrating the pAH68-*frt* integration plasmid into TB357 using the pAH69 helper plasmid. pAH68-*frt*-cat is a CRIM-based plasmid (Haldimann & Wanner, 2001) previously constructed in our laboratory that integrates into the HK022 attachment site and contains a chloramphenicol resistance marker.  $\lambda$  **kan** was constructed by recombineering (Oppenheim et al., 2004). Specifically,  $\lambda$  PaPa lysogens were first transformed with pKD46 (Datsenko & Wanner, 2000). Arabinose-induced electro-competent cells were then electroporated with the  $\Delta bor::kanR$  PCR product, followed by selection for kanamycin resistant colonies. The ability of the resulting clones to release free phage conferring kanamycin resistance upon lysogenization was confirmed. The  $\Delta bor::kanR$  PCR fragment was obtained by PCR using the *lambda\_bor\_kan\_fwd* and *lambda\_bor\_kan\_rev* primers and pKD4 as the template (Datsenko & Wanner, 2000). The EcoRII RM system was amplified from pR209 using primers *fw\_NheI\_EcoRII* and *rv\_SalI\_EcoRII*. The resulting fragment was cut with NheI and SalI and cloned into pBR322 via the corresponding restriction sites, resulting in **pMPR001**.

#### 4.5.15 Bacterial strains, plasmids and phages

Name	Genotype	Source, reference
DH5 $\alpha$	<i>F</i> <sup>-</sup> , $\lambda$ <sup>-</sup> , $\Phi$ 80 <i>lacZ</i> $\Delta$ <i>M15</i> , $\Delta$ ( <i>lacZYA-argF</i> ), <i>U169</i> , <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> ( <i>rK</i> <sup>-</sup> , <i>mK</i> <sup>+</sup> ), <i>phoA supE44</i> , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i>	Lab collection
DH5 $\alpha$ $\lambda$ <i>pir</i> <sup>+</sup>	<i>F</i> <sup>-</sup> , $\Phi$ 80 <i>lacZ</i> $\Delta$ <i>M15</i> , $\Delta$ ( <i>lacZYA-argF</i> ), <i>U169</i> , <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> ( <i>rK</i> <sup>-</sup> , <i>mK</i> <sup>+</sup> ), <i>phoA supE44</i> , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i> , $\lambda$ <i>pir</i> <sup>+</sup>	Lab collection
MG1655	<i>F</i> <sup>-</sup> , $\lambda$ <sup>-</sup> , <i>ilvG</i> <sup>-</sup> , <i>rfb-50</i> , <i>rph-1</i>	Lab collection
MP084	MG1655, $\Delta$ <i>mcrB</i>	This work
MP085	MG1655, $\Delta$ <i>attHK022::cat</i> , $\Delta$ <i>araA</i>	This work
TB357	MG1655, $\Delta$ <i>araA</i>	This work
C-1	<i>E. coli</i> C	CGSC
WA251	<i>E. coli</i> B, <i>lon</i> <sup>-</sup> , <i>mal</i> <sup>+</sup>	CGSC
NK354	$\Delta$ <i>hsd</i> <sub><i>EcoKI</i></sub> <i>hsd</i> <sup>+</sup> <sub><i>EcoAI</i></sub>	(Makovets et al., 2004)
NK402	$\Delta$ <i>hsd</i> <sub><i>EcoKI</i></sub> , <i>lac::(hsd</i> <sup>+</sup> <sub><i>EcoR124I</i></sub> , <i>cat)</i>	(Makovets et al., 2004)
pBR322 $\Delta$ P <sub>tet</sub>	<i>pMB1 ori</i> , <i>bla</i> , $\Delta$ P <sub>tet</sub>	(Pleška et al., 2016)
pBR322 $\Delta$ P <sub>tet</sub> EcoRI (R+M+)	<i>pMB1 ori</i> , <i>bla</i> , $\Delta$ P <sub>tet</sub> , EcoRI (R+M+)	(Pleška et al., 2016)
pBR322 $\Delta$ P <sub>tet</sub> EcoRV (R+M+)	<i>pMB1 ori</i> , <i>bla</i> , $\Delta$ P <sub>tet</sub> , EcoRV (R+M+)	(Pleška et al., 2016)
pR209	<i>p15A ori</i> , <i>bla</i> , EcoRII (R+M+)	(Bhagwat & Johnson, 1990)
pMPR001	<i>pMB1 ori</i> , <i>bla</i> , EcoRII (R+M+)	This work
pPvuII 3.4	<i>pMB1 ori</i> , <i>bla</i> , PvuII (R+M+)	(Blumenthal et al., 1985)
pRR0	<i>pMB1 ori</i> , <i>bla</i> , EcoVIII (R+M+)	Iwona Mruk
pNR201	<i>p15A ori</i> , <i>cat</i> , EcoP1I	(Hümbelin et al., 1988)
pNR301	<i>p15A ori</i> , <i>cat</i> , EcoP15I	(Hümbelin et al., 1988)
pAH68- <i>frt</i> -Cam <sup>R</sup>	<i>R6K ori</i> , <i>cat -frt</i> , <i>attPHK022</i>	Lab collection
$\lambda$ vir	Virulent mutant of Phage $\lambda$	Sylvain Moineau
$\lambda$ kan	<i>kan</i>	This work

#### 4.5.16 List of primers

Name	Sequence (5' to 3')
lambda_bor_kan_fwd	TACGATTCTGCGAACTTCAAAAAGCATCGGGAATAACACCTCA CGCTGCCGCAAGCACTC
lambda_bor_kan_rev	TGCAGATAGAGTTGCCCATATCGATGGGCAACTCATGCAAAGC GCTTTTGAAGCTGGGGTG
5_araA_KO	GCTGCCCAGGCCGTTGCGACTCTATAAGGACACGATAATGATT CCGGGGATCCGTCGACC
3_araA_KO	GTCAGCGTCGCATCAGGCGTTACATACCGGATGCGGCTACTGT AGGCTGGAGCTGCTTC
fw_NheI_EcoRII	AATTAAgctagcATCCCACAACCTCATGAGCC
rv_SalI_EcoRII	AATTAAgtcgacGCTCAACCACCATTTCGCAG

## 5 References

- Abedon, S. T. (2012). Bacterial “immunity” against bacteriophages. *Bacteriophage*, 2(1), 50–54.
- Akerlund, T., Nordström, K., & Bernander, R. (1995). Analysis of cell size and DNA content in exponentially growing and stationary-phase batch cultures of *Escherichia coli*. *Journal of Bacteriology*, 177(23), 6791–6797.
- Arber, W. (2000). Genetic variation: Molecular mechanisms and impact on microbial evolution. *FEMS Microbiology Reviews*, 24(1), 1–7.
- Arber, W., & Dussoix, D. (1962). Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *Journal of Molecular Biology*, 5(1), 18–36.
- Arber, W., & Linn, S. (1969). DNA Modification and Restriction. *Annual Review of Biochemistry*, 38(1), 467–500.
- Asakura, Y., & Kobayashi, I. (2009). From damaged genome to cell surface: transcriptome changes during bacterial cell death triggered by loss of a restriction-modification gene complex. *Nucleic Acids Research*, 37(9), 3021–31.
- Asakura, Y., Kojima, H., & Kobayashi, I. (2011). Evolutionary genome engineering using a restriction-modification system. *Nucleic Acids Research*, 39(20), 9034–9046.
- Avery, S. V. (2006). Microbial cell individuality and the underlying sources of heterogeneity. *Nature Reviews Microbiology*, 4(8), 577–587.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., ... Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2, 2006.0008.
- Barondess, J. J., & Beckwith, J. (1990). A bacterial virulence determinant encoded by lysogenic coliphage  $\lambda$ . *Nature*, 346(6287), 871–874.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., ... Horvath, P.

- (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* (New York, N.Y.), 315(5819), 1709–12.
- Berkner, K. L., & Folk, W. R. (1983). An assay for the rates of cleavage of specific sites in DNA by restriction endonucleases: Its use to study the cleavage of phage  $\lambda$  DNA by EcoRI and phage P22 DNA containing thymine or 5-bromouracil by HindIII. *Analytical Biochemistry*, 129(2), 446–456.
- Berngruber, T. W., Lion, S., & Gandon, S. (2013). Evolution of suicide as a defence strategy against pathogens in a spatially structured environment. *Ecology Letters*, 16(4), 446–453.
- Bertani, G., & Weigle, J. (1953). Host Controlled Variation In Bacterial Viruses. *Journal of Bacteriology*, 65(2), 113–121.
- Betlach, M., Hershfield, V., Chow, L., Brown, W., Goodman, H., & Boyer, H. W. (1976). A restriction endonuclease analysis of the bacterial plasmid controlling the *ecoRI* restriction and modification of DNA. *Federation Proceedings*, 35(9), 2037–2043.
- Bhagwat, A., & Johnson, B. (1990). Primary sequence of the *EcoRII* endonuclease and properties of its fusions with beta-galactosidase. *Journal of Biological Chemistry*, 265(2), 767–773.
- Bickle, T. a, & Krüger, D. H. (1993). Biology of DNA restriction. *Microbiological Reviews*, 57(2), 434–50.
- Blumenthal, R. M., & Cheng, X. (2002). Restriction-Modification Systems. In *Modern Microbial Genetics* (pp. 177–225). New York, USA: John Wiley & Sons, Inc.
- Blumenthal, R. M., Gregory, S. a, & Cooperider, J. S. (1985). Cloning of a restriction-modification system from *Proteus vulgaris* and its use in analyzing a methylase-sensitive phenotype in *Escherichia coli*. *Journal of Bacteriology*, 164(2), 501–9.
- Bobay, L.-M., Rocha, E. P. C., & Touchon, M. (2013). The Adaptation of Temperate Bacteriophages to Their Host Genomes. *Molecular Biology and Evolution*, 30(4), 737–751.

- Bondy-Denomy, J., Qian, J., Westra, E. R., Buckling, A., Guttman, D. S., Davidson, A. R., & Maxwell, K. L. (2016). Prophages mediate defense against phage infection through diverse mechanisms. *The ISME Journal*, 22, 1–13.
- Boots, M. (2011). The evolution of resistance to a parasite is determined by resources. *The American Naturalist*, 178(2), 214–220.
- Brown, S. P., Le Chat, L., De Paepe, M., & Taddei, F. (2006). Ecology of Microbial Invasions: Amplification Allows Virus Carriers to Invade More Rapidly When Rare. *Current Biology*, 16(20), 2048–2052.
- Brüssow, H., Canchaya, C., Hardt, W., & Bru, H. (2004). Phages and the Evolution of Bacterial Pathogens : from Genomic Rearrangements to Lysogenic Conversion Phages and the Evolution of Bacterial Pathogens : from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews*, 68(3), 560–602.
- Canchaya, C., Fournous, G., & Brüssow, H. (2004). The impact of prophages on bacterial chromosomes. *Molecular Microbiology*, 53(1), 9–18.
- Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L., & Brüssow, H. (2003). Phage as agents of lateral gene transfer. *Current Opinion in Microbiology*, 6(4), 417–424.
- Casjens, S. (2003). Prophages and bacterial genomics: What have we learned so far? *Molecular Microbiology*, 49(2), 277–300.
- Chait, R., Shrestha, S., Shah, A. K., Michel, J.-B., & Kishony, R. (2010). A Differential Drug Screen for Compounds That Select Against Antibiotic Resistance. *PLoS ONE*, 5(12), e15179.
- Chandrasegaran, S., & Smith, H. (1988). Amino acid sequence homologies among twenty-five restriction endonucleases and methylases. *Structure & Expression*.
- Chang, S., & Cohen, S. N. (1977). In vivo site-specific genetic recombination promoted by the EcoRI restriction endonuclease. *Proceedings of the National Academy of Sciences of the United States of America*, 74(11), 4811–4815.

- Chiang, C. S., & Bremer, H. (1988). Stability of pBR322-derived plasmids. *Plasmid*, 20(3), 207–220.
- Chopin, M. C., Chopin, A., & Bidnenko, E. (2005). Phage abortive infection in lactococci: Variations on a theme. *Current Opinion in Microbiology*, 8(4), 473–479.
- Clokier, M. R., Millard, A. D., Letarov, A. V., & Heaphy, S. (2011). Phages in nature. *Bacteriophage*, 1(1), 31–45.
- Corvaglia, A. R., Francois, P., Hernandez, D., Perron, K., Linder, P., & Schrenzel, J. (2010). A type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical *Staphylococcus aureus* strains. *Proceedings of the National Academy of Sciences*, 107(26), 11954–11958.
- Cox, M. M., Layton, S. L., Jiang, T., Cole, K., Hargis, B. M., Berghman, L. R., ... Kwon, Y. M. (2007). Scarless and site-directed mutagenesis in *Salmonella enteritidis* chromosome. *BMC Biotechnology*, 7, 59.
- Cromie, G. A., & Leach, D. R. F. (2001). Recombinational repair of chromosomal DNA double-strand breaks generated by a restriction endonuclease. *Molecular Microbiology*, 41(4), 873–883.
- Darmon, E., Eykelenboom, J. K., Lopez-Vernaza, M. A., White, M. A., & Leach, D. R. F. (2014). Repair on the Go: *E. coli* Maintains a High Proliferation Rate while Repairing a Chronic DNA Double-Strand Break. *PLoS ONE*, 9(10), e110784.
- Datsenko, K. a., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 6640–6645.
- Davies, E. V., James, C. E., Williams, D., O'Brien, S., Fothergill, J. L., Haldenby, S., ... Brockhurst, M. A. (2016). Temperate phages both mediate and drive adaptive evolution in pathogen biofilms. *Proceedings of the National Academy of Sciences*, 113(29), 8266–8271.
- Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., ... Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*.

- Journal of Bacteriology, 190(4), 1390–1400.
- Doublet, B., Douard, G., Targant, H., Meunier, D., Madec, J. Y., & Cloeckert, A. (2008). Antibiotic marker modifications of  $\lambda$  Red and FLP helper plasmids, pKD46 and pCP20, for inactivation of chromosomal genes using PCR products in multidrug-resistant strains. *Journal of Microbiological Methods*, 75(2), 359–361.
- Dykhuizen, D. E., & Hartl, D. L. (1983). Selection in chemostats. *Microbiological Reviews*, 47(2), 150–168.
- Edgar, R., & Qimron, U. (2010). The *Escherichia coli* CRISPR system protects from  $\lambda$  lysogenization, lysogens, and prophage induction. *Journal of Bacteriology*, 192(23), 6291–4.
- Edlin, G., Lin, L., & Bitner, R. (1977). Reproductive fitness of P1, P2, and Mu lysogens of *Escherichia coli*. *Journal of Virology*, 21(2), 560–4.
- Elhai, J. (2001). Determination of bias in the relative abundance of oligonucleotides in DNA sequences. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, 8(2), 151–175.
- Ellis, E. L., & Delbrück, M. (1939). The Growth of Bacteriophage. *The Journal of General Physiology*, 22(3), 365–84.
- Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science (New York, N.Y.)*, 297(5584), 1183–1186.
- Enikeeva, F. N., Severinov, K. V., & Gelfand, M. S. (2010). Restriction–modification systems and bacteriophage invasion: Who wins? *Journal of Theoretical Biology*, 266(4), 550–559.
- Erez, Z., Steinberger-Levy, I., Shamir, M., Doron, S., Stokar-Avihail, A., Peleg, Y., ... Sorek, R. (2017). Communication between viruses guides lysis–lysogeny decisions. *Nature*, 541(7638), 488–493.
- Fogg, P. C. M., Allison, H. E., Saunders, J. R., & McCarthy, A. J. (2010). Bacteriophage lambda: a paradigm revisited. *Journal of Virology*, 84(13), 6876–9.
- Friedberg, E. C., Walker, G. C., Siede, W., & Wood, R. D. (2005). DNA repair and mutagenesis.

American Society for Microbiology Press.

- Furuta, Y., & Kobayashi, I. (2012). Restriction-Modification Systems as Mobile Epigenetic Elements. In A. Roberts & P. Mullany (Eds.), *Bacterial Integrative Mobile Genetic Elements* (pp. 1–19). Landes Bioscience.
- Gandon, S. (2016). Why Be Temperate: Lessons from Bacteriophage  $\lambda$ . *Trends in Microbiology*, 24(5), 356–365.
- Gelfand, M. S., & Koonin, E. V. (1997). Avoidance of palindromic words in bacterial and archaeal genomes: A close connection with restriction enzymes. *Nucleic Acids Research*, 25(12), 2430–2439.
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 93(5), 2340–2361.
- Goldberg, G. W., Jiang, W., Bikard, D., & Marraffini, L. a. (2014). Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature*, 514(7524), 633–637.
- Goldberg, G. W., & Marraffini, L. A. (2015). Resistance and tolerance to foreign elements by prokaryotic immune systems — curating the genome. *Nature Reviews Immunology*, 15(11), 717–724.
- Gómez, P., & Buckling, A. (2011). Bacteria-phage antagonistic coevolution in soil. *Science* (New York, N.Y.), 332(6025), 106–109.
- Haldimann, A., & Wanner, B. L. (2001). Conditional-Replication , Integration , Excision , and Retrieval Plasmid-Host Systems for Gene Structure-Function Studies of Bacteria. *Journal of Bacteriology*, 183, 6384–6393.
- Handa, N., Ichige, A., Kusano, K., & Kobayashi, I. (2000). Cellular responses to postsegregational killing by restriction-modification genes. *Journal of Bacteriology*, 182(8), 2218–2229.
- Handa, N., & Kobayashi, I. (1999). Post-segregational killing by restriction modification gene complexes: observations of individual cell deaths. *Biochimie*, 81(8–9), 931–8.

- Hayes, F. (2003). Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science (New York, N.Y.)*, 301(5639), 1496–9.
- Heitman, J., Ivanenko, T., & Kiss, A. (1999). DNA nicks inflicted by restriction endonucleases are repaired by a RecA- and RecB-dependent pathway in *Escherichia coli*. *Molecular Microbiology*, 33(6), 1141–1151.
- Heitman, J., Zinder, N. D., & Model, P. (1989). Repair of the *Escherichia coli* chromosome after in vivo scission by the EcoRI endonuclease. *Proceedings of the National Academy of Sciences of the United States of America*, 86(7), 2281–5.
- Hong, J.-S., Smith, G. R., & Ames, B. N. (1971). Adenosine 3':5'-Cyclic Monophosphate Concentration in the Bacterial Host Regulates the Viral Decision between Lysogeny and Lysis. *Proceedings of the National Academy of Sciences of the United States of America*, 68(9), 2258–2262.
- Houte, S. van, Buckling, A., & Westra, E. R. (2016). Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiology and Molecular Biology Reviews*, 80(3), 745–763.
- Houte, S. van, Ekroth, A. K. E., Broniewski, J. M., Chabas, H., Ashby, B., Gandon, S., ... Westra, E. R. (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature*, 532(7599), 385–388.
- Huh, D., & Paulsson, J. (2011). Non-genetic heterogeneity from stochastic partitioning at cell division. *Nature Genetics*, 43(2), 95–100.
- Hümbelin, M., Suri, B., Rao, D. N., Hornby, D. P., Eberle, H., Pripfl, T., ... Bickle, T. A. (1988). Type III DNA restriction and modification systems EcoP1 and EcoP15. Nucleotide sequence of the EcoP1 operon, the EcoP15 mod gene and some EcoP1 mod mutants. *Journal of Molecular Biology*, 200(1), 23–9.
- Ichige, A., & Kobayashi, I. (2005). Stability of EcoRI restriction-modification enzymes in vivo differentiates the EcoRI restriction-modification system from other postsegregational cell killing systems. *Journal of Bacteriology*, 187(19), 6612–6621.
- Jeltsch, A. (2003). Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? *Gene*, 317, 13–16.

- Karlin, S., Burge, C., & Campbell, A. M. (1992). Statistical analyses of counts and distributions of restriction sites in DNA sequences. *Nucleic Acids Research*, 20(6), 1363–1370.
- Knowles, B., Bailey, B., Boling, L., Breitbart, M., Cobián-Güemes, A., del Campo, J., ... Rohwer, F. (2017). Variability and host density independence in inductions-based estimates of environmental lysogeny. *Nature Microbiology*, 2(0), 17064.
- Knowles, B., Silveira, C. B., Bailey, B. a, Barott, K., Cantu, V. a, Cobián-Güemes, a G., ... Rohwer, F. (2016). Lytic to temperate switching of viral communities. *Nature*, 531(7595), 466–70.
- Kobayashi, I. (2001). Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Research*, 29(18), 3742–3756.
- Korona, R., Korona, B., & Levin, B. R. (1993). Sensitivity of naturally occurring coliphages to type I and type II restriction and modification. *Journal of General Microbiology*, 139(1993), 1283–1290.
- Korona, R., & Levin, B. (1993). Phage-Mediated Selection and the Evolution and Maintenance of Restriction-Modification. *Evolution*, 47(2), 556–575.
- Koskella, B., & Brockhurst, M. A. (2014). Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews*, 38(5), 916–931.
- Kourilsky, P. (1973). Lysogenization by bacteriophage lambda. *MGG Molecular & General Genetics*, 122(2), 183–195.
- Krüger, D. H., & Bickle, T. a. (1983). Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiological Reviews*, 47(3), 345–360.
- Krüger, D. H., Kupper, D., Meisel, A., Reuter, M., & Schroeder, C. (1995). The significance of distance and orientation of restriction endonuclease recognition sites in viral DNA genomes. *FEMS Microbiology Reviews*, 17(1–2), 177–184.
- Kusano, K., Naito, T., Handa, N., & Kobayashi, I. (1995). Restriction-modification systems as

- genomic parasites in competition for specific sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 92(24), 11095–9.
- Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews. Microbiology*, 8(5), 317–27.
- Lenski, R. E. (1988). Experimental studies of pleiotropy and epistasis in *Escherichia coli*. I. variation in competitive fitness among mutants resistant to virus T4. *Evolution*.
- Lenski, R. E., & Levin, B. R. (1985). Constraints on the Coevolution of Bacteria and Virulent Phage: A Model, Some Experiments, and Predictions for Natural Communities, 125(4), 585–602.
- Levin, B. R., Moineau, S., Bushman, M., & Barrangou, R. (2013). The Population and Evolutionary Dynamics of Phage and Bacteria with CRISPR–Mediated Immunity. *PLoS Genetics*, 9(3), e1003312.
- Liang, J., & Blumenthal, R. M. (2013). Naturally-occurring, dually-functional fusions between restriction endonucleases and regulatory proteins. *BMC Evolutionary Biology*, 13(1), 218.
- Lieb, M. (1953). The establishment of lysogenicity in *Escherichia coli*. *Journal of Bacteriology*, 65(6), 642–51.
- Lin, L., Bitner, R., & Edlin, G. (1977). Increased reproductive fitness of *Escherichia coli* lambda lysogens. *Journal of Virology*, 21(2), 554–559.
- Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. a., & Wilson, G. G. (2014). Type i restriction enzymes and their relatives. *Nucleic Acids Research*, 42(1), 20–44.
- Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. a., Wilson, G. G., & Murray, N. E. (2014). Highlights of the DNA cutters: A short history of the restriction enzymes. *Nucleic Acids Research*, 42(1), 3–19.
- Loenen, W. A. M., & Raleigh, E. A. (2014). The other face of restriction: Modification-dependent enzymes. *Nucleic Acids Research*, 42(1), 56–69.
- Luria, S. E., & Human, M. L. (1952). A nonhereditary, host-induced variation of bacterial

- viruses. *Journal of Bacteriology*, 64(4), 557–69.
- Lutz, R., & Bujard, H. (1997). Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research*, 25(6), 1203–1210.
- Lwoff, A. (1953). Lysogeny. *Bacteriological Reviews*, 17(4), 269–337.
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., ... Koonin, E. V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*, 13(11), 722–736.
- Makovets, S., Powell, L. M., Titheradge, A. J. B., Blakely, G. W., & Murray, N. E. (2004, November 11). Is modification sufficient to protect a bacterial chromosome from a resident restriction endonuclease? *Molecular Microbiology*.
- Marraffini, L. A., & Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, 463(7280), 568–571.
- McAdams, H. H., & Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 94(3), 814–819.
- McAdams, H. H., & Arkin, A. (1999). It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in Genetics*, 15(2), 65–69.
- Meisel, A., Bickle, T. A., Krieger, D. H., & Schroeder, C. (1992). Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature*, 355(6359), 467–469.
- Mruk, I., & Blumenthal, R. M. (2008). Real-time kinetics of restriction-modification gene expression after entry into a new host cell. *Nucleic Acids Research*, 36(8), 2581–93.
- Mruk, I., & Kobayashi, I. (2014, August 13). To be or not to be: Regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Research*.
- Mruk, I., Liu, Y., Ge, L., & Kobayashi, I. (2011). Antisense RNA associated with biological regulation of a restriction-modification system. *Nucleic Acids Research*, 39(13), 5622–5632.

- Murray, N. E. (2002). Immigration control of DNA in bacteria: self versus non-self. *Microbiology*, 148(1), 3–20.
- Murray, N. E., & Murray, K. (1974). Manipulation of restriction targets in phage  $\lambda$  to form receptor chromosomes for DNA fragments. *Nature*, 251(5475), 476–481.
- Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K., & Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology*, 20(1), 87–90.
- Nagornykh, M. O., Bogdanova, E. S., Protsenko, A. S., Solonin, A. S., Zakharova, M. V., & Severinov, K. V. (2008). Regulation of gene expression in a type II restriction-modification system. *Russian Journal of Genetics*.
- Naito, T., Kusano, K., & Kobayashi, I. (1995). Selfish behavior of restriction-modification systems. *Science (New York, N.Y.)*, 267(5199), 897–899.
- Nakayama, Y., & Kobayashi, I. (1998). Restriction-modification gene complexes as selfish gene entities: roles of a regulatory system in their establishment, maintenance, and apoptotic mutual exclusion. *Proceedings of the National Academy of Sciences of the United States of America*, 95(11), 6442–7.
- Nölling, J., & Vos, W. De. (1992). Characterization of the Archaeal , Plasmid-Encoded Type II Restriction-Modification System MthTI from *Methanobacterium thermoformicum* THF : homology to the bacterial NgoPII system from *Neisseria gonorrhoeae*. *Journal of Bacteriology*, 174(17), 5719–5726.
- O'Brien, A. D., Newland, J. W., Miller, S. F., Holmes, R. K., Smith, H. W., & Formal, S. B. (1984). Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science (New York, N.Y.)*, 226(4675), 694–6.
- O'Neill, M., Chen, a, & Murray, N. E. (1997). The restriction-modification genes of *Escherichia coli* K-12 may not be selfish: they do not resist loss and are readily replaced by alleles conferring different specificities. *Proceedings of the National Academy of Sciences of the United States of America*, 94(26), 14596–601.
- Obeng, N., Pratama, A. A., & Elsas, J. D. van. (2016). The Significance of Mutualistic Phages

- for Bacterial Ecology and Evolution. *Trends in Microbiology*, 24(6), 440–449.
- Ochman, H., Lawrence, J. G., & Groisman, E. a. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*, 405(6784), 299–304.
- Ohshima, Y., Schumacher-Perdreau, F., Peters, G., & Pulverer, G. (1988). The role of capsule as a barrier to bacteriophage adsorption in an encapsulated *Staphylococcus simulans* strain. *Medical Microbiology and Immunology*, 177(4), 229–233.
- Oliveira, P. H., Touchon, M., & Rocha, E. P. C. (2014). The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Research*, 42(16), 10618–10631.
- Oliveira, P. H., Touchon, M., & Rocha, E. P. C. (2016). Regulation of genetic flux between bacteria by restriction–modification systems. *Proceedings of the National Academy of Sciences*, 113(20), 5658–5663.
- Oliver, K., Degnan, P., Hunter, M., & Moran, N. (2009). Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science (New York, N.Y.)*, 325(5943), 992–994.
- Oppenheim, A. B., Rattray, A. J., Bubunencko, M., Thomason, L. C., & Court, D. L. (2004). In vivo recombineering of bacteriophage  $\lambda$  by PCR fragments and single-strand oligonucleotides. *Virology*, 319(2), 185–189.
- Oppenheim, A. B., & Salomon, D. (1972). Studies on partially virulent mutants of lambda bacteriophage - II. The mechanism of overcoming repression. *MGG Molecular & General Genetics*, 115(2), 101–114.
- Parma, D. H., Snyder, M., Sobolevski, S., Nawroz, M., Brody, E., & Gold, L. (1992). The Rex System of Bacteriophage-Lambda - Tolerance and Altruistic Cell-Death. *Genes & Development*, 6(3), 497–510.
- Pennington, J. M., & Rosenberg, S. M. (2007). Spontaneous DNA breakage in single living *Escherichia coli* cells. *Nature Genetics*, 39(6), 797–802.
- Pingoud, V., Kubareva, E., Stengel, G., Friedhoff, P., Bujnicki, J. M., Urbanke, C., ... Pingoud,

- A. (2002). Evolutionary relationship between different subgroups of restriction endonucleases. *The Journal of Biological Chemistry*, 277(16), 14306–14.
- Pleška, M., Qian, L., Okura, R., Bergmiller, T., Wakamoto, Y., Kussell, E., & Guet, C. C. (2016). Bacterial autoimmunity due to a restriction-modification system. *Current Biology*, 26(3), 404–409.
- Qian, L., & Kussell, E. (2012). Evolutionary Dynamics of Restriction Site Avoidance. *Physical Review Letters*, 108(15), 158105.
- Ramsey, S., Orrell, D., Bolouri, H., & Others. (2005). Dizzy: stochastic simulation of large-scale genetic regulatory networks. *Journal of Bioinformatics and Computational Biology*, 3(2), 415–436.
- Rao, D. N., Dryden, D. T. F., & Bheemanaik, S. (2014). Type III restriction-modification enzymes: A historical perspective. *Nucleic Acids Research*, 42(1), 45–55.
- Rau, D. C., & Sidorova, N. Y. (2010). Diffusion of the Restriction Nuclease EcoRI along DNA. *Journal of Molecular Biology*, 395(2), 408–416.
- Refardt, D., Bergmiller, T., & Kümmerli, R. (2013). Altruism can evolve when relatedness is low: evidence from bacteria committing suicide upon phage infection. *Proceedings. Biological Sciences / The Royal Society*, 280(1759), 20123035.
- Rice, S. A., Tan, C. H., Mikkelsen, P. J., Kung, V., Woo, J., Tay, M., ... Kjelleberg, S. (2009). The biofilm life cycle and virulence of *Pseudomonas aeruginosa* are dependent on a filamentous prophage. *The ISME Journal*, 3(3), 271–282.
- Ripp, S., & Miller, R. V. (1997). The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. *Microbiology*, 143(6), 2065–2070.
- Roberts, R. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research*, 31(7), 1805–1812.
- Roberts, R., Vincze, T., Posfai, J., & Macelis, D. (2015). REBASE-A database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Research*, 38(SUPPL.1), 2014–2015.

- Rocha, E. P. C., Danchin, A., & Viari, A. (2001). Evolutionary Role of Restriction- Modification Systems as Revealed by Comparative Genome Analysis. *Genome Research*, 11(6), 946–958.
- Sargentini, N. J., Diver, W. P., & Smith, K. C. (1983). The effect of growth conditions on inducible, recA-dependent resistance to X rays in *Escherichia coli*. *Radiat Res*, 93(2), 364–380.
- Semenova, E., Minakhin, L., Bogdanova, E., Nagornykh, M., Vasilov, A., Heyduk, T., ... Severinov, K. (2005). Transcription regulation of the EcoRV restriction-modification system. *Nucleic Acids Research*, 33(21), 6942–51.
- St-Pierre, F., & Endy, D. (2008). Determination of cell fate selection during phage lambda infection. *Proceedings of the National Academy of Sciences of the United States of America*, 105(52), 20705–20710.
- Stewart, F. M., & Levin, B. R. (1984). The population biology of bacterial viruses: why be temperate. *Theoretical Population Biology*, 26(1), 93–117.
- Takahashi, N., Naito, Y., Handa, N., & Kobayashi, I. (2002). A DNA methyltransferase can protect the genome from postdisturbance attack by a restriction-modification gene complex. *Journal of Bacteriology*, 184(22), 6100–6108.
- Takahashi, N., Ohashi, S., Sadykov, M. R., Mizutani-Ui, Y., & Kobayashi, I. (2011). IS-linked movement of a restriction-modification system. *PloS One*, 6(1), e16554.
- Thomason, L. C., Sawitzke, J. A., Li, X., Costantino, N., & Court, D. L. (2014). Specialized techniques: Recombineering: Genetic engineering in bacteria using homologous recombination. *Current Protocols in Molecular Biology*, 1(SUPL.106), 1.
- Tock, M. R., & Dryden, D. T. F. (2005). The biology of restriction and anti-restriction. *Current Opinion in Microbiology*, 8(4), 466–72.
- Touchon, M., Bernheim, A., & Rocha, E. P. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *The ISME Journal*, 10(11), 1–11.
- Vale, P. F., Lafforgue, G., Gatchitch, F., Gardan, R., Moineau, S., & Gandon, S. (2015). Costs

- of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. *Proceedings of the Royal Society B: Biological Sciences*, 282(1812), 20151270.
- Vasu, K., Nagamalleswari, E., & Nagaraja, V. (2012). Promiscuous restriction is a cellular defense strategy that confers fitness advantage to bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 109(20), E1287-93.
- Vasu, K., & Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiology and Molecular Biology Reviews : MMBR*, 77(1), 53–72.
- Waldor, M., & Mekalanos, J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science (New York, N.Y.)*, 272(5270), 1910–1914.
- Wang, P., Robert, L., Pelletier, J., Dang, W. L., Taddei, F., Wright, A., & Jun, S. (2010). Robust growth of *Escherichia coli*. *Current Biology*, 20(12), 1099–1103.
- Wang, X., Kim, Y., Ma, Q., Hong, S. H., Pokusaeva, K., Sturino, J. M., & Wood, T. K. (2010). Cryptic prophages help bacteria cope with adverse environments. *Nature Communications*, 1(9), 147.
- Westra, E. R., van Houte, S., Oyesiku-Blakemore, S., Makin, B., Broniewski, J. M., Best, A., ... Buckling, A. (2015). Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense. *Current Biology*, 25(8), 1043–1049.
- Wigington, C. H., Sonderegger, D., Brussaard, C. P. D., Buchan, A., Finke, J. F., Fuhrman, J. A., ... Weitz, J. S. (2016). Re-examination of the relationship between marine virus and microbial cell abundances. *Nature Microbiology*, 1(3), 15024.
- Wilson, G. G., & Murray, N. E. (1991). Restriction and modification systems. *Annual Review of Genetics*, 25, 585–627.
- Zeng, L., Skinner, S. O., Zong, C., Sippy, J., Feiss, M., & Golding, I. (2010). Decision making at a subcellular level determines the outcome of bacteriophage infection. *Cell*, 141(4), 682–91.