2023

Recombination and Excision: DNA Repair Proteins in Prokaryotic Host-Virus Conflicts

Amer Azim Hossain

Recombination and excision: DNA repair proteins in prokaryotic host-virus conflicts

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements

for the degree of Doctor of Philosophy

By

Amer Azim Hossain

December 2023

RECOMBINATION AND EXCISION: DNA REPAIR PROTEINS IN PROKARYOTIC HOST-VIRUS CONFLICTS

Amer Azim Hossain, Ph.D.
The Rockefeller University 2023

Bacteriophages, or simply phages, are viruses that infect bacteria. They are the most abundant biological entity on our planet and outnumber bacteria 10:1 in the ocean. In response to this threat, bacteria have evolved a diverse battery of immune systems that prevent infection, which in turn has resulted in the development of numerous counter-defense mechanisms by phages. This evolutionary arms race drives molecular innovations and presents exciting avenues for the discovery of new molecular biology and new biotechnology tools, such as restriction enzymes and CRISPR-Cas9. My thesis investigates how mechanisms of DNA repair, specifically recombination and base excision, have been co-opted by phages and bacteria to execute non-canonical immune and counter-immune functions in prokaryotic host-virus genetic conflicts.

CRISPR-Cas adaptive immune systems, found in nearly half of all bacteria, use sequence-specific guide RNAs to cleave the genetic material of infecting phages. Bacteria and some phages encode recombination systems that could repair the cleaved viral DNA. At the outset of my PhD thesis, it was unknown whether phages could counteract CRISPR-Cas cleavage of phage DNA by repairing CRISPR-induced

DNA breaks. Bacteriophage λ, which infects *Escherichia coli*, encodes the Red system (*gam-exo-bet*) to promote recombination between related phages. Here, using molecular genetics and sequencing, I show that λ Red mediates evasion of CRISPR-Cas targeting in *E. coli*. Gam inhibits the host *E. coli* RecBCD recombination system, allowing recombination and repair of the cleaved DNA by the phage Exo-Beta exonuclease-recombinase. Repair by Exo-Beta promotes mutations, deletions, and genomic rearrangements within the target sequence in phage DNA to prevent recognition by CRISPR. I find that λ Red recombination is strikingly more efficient than the host's RecBCD-RecA recombination pathway in the production of large numbers of phages that escape CRISPR targeting. These findings establish recombination-mediated DNA repair as a novel viral "anti-CRISPR" strategy that, rather than binding CRISPR-Cas nucleases and impeding their activity, provides a solution to evade the CRISPR-Cas immune response *after* it has been set off. While recombinases canonically function in DNA repair, my findings reveal an additional role for Red-like recombination systems in countering bacterial immunity, through the protection of phages against sequence-specific nucleases. Based on these findings, I speculate that the counter-immune advantage imparted by Red-like systems may facilitate their spread across bacteriophage genomes.

For the second half of my thesis, I set out to discover novel defense systems in bacteria, specifically those that target viral DNA. To achieve this, I pioneered a new screening methodology to discover anti-phage defense systems from unculturable

microorganisms using diverse bacterial metagenomic DNA libraries. These metagenomic libraries contain millions of DNA sequences from different microorganisms that are absent in available genetic databases. While bioinformatic mining has led to the discovery of many new bacterial immune systems, the genetic screening of DNA libraries has the advantage of: (i) examining unsequenced DNA, including the “dark matter” of microbial genomes, and (ii) discovering novel defense systems that cannot be predicted via computational analyses. By subjecting the metagenomic libraries (cloned in *E. coli*) to phage infection and isolating resistant colonies, I discovered a novel bacterial DNA glycosylase that I named Brig1 (bacteriophage replication inhibition DNA glycosylase 1). Brig1 provides immunity against phages that carry “hypermodified” DNA nucleobases, specifically alpha-glucosyl-hydroxymethylcytosine nucleobases. Brig1 excises these nucleobases from the genome of T-even phages, such as coliphage T4, to generate abasic sites that inhibit DNA replication, which constitutes a novel anti-phage defense mechanism. Many phages have evolved to introduce DNA modifications to avoid recognition and cleavage by bacterial nucleases, including CRISPR-Cas and restriction endonucleases. Brig1 supplies the next step on the bacterial side of the arms race, reestablishing the restriction of phages with modified genomes. Structural predictions suggest that Brig1 is part of a novel family of bacterial anti-phage DNA glycosylases that evolved from uracil DNA glycosylases involved in base excision repair, a pathway that removes mis-incorporated uracil bases from DNA. Interestingly, Brig1 homologs are present in multiple phage defense loci across distinct clades of bacteria, and these will be the focus of future studies. Future work will also employ

the same metagenomic screening approach, but infecting with phages harboring different DNA modifications, to drive the discovery of new DNA glycosylases, restriction enzymes and DNA repair modules that target modified DNA.

Overall, my thesis establishes DNA repair proteins as emerging players in prokaryotic host-virus warfare and will spur future work that explores the co-option of these proteins in varying immune and counter-immune contexts in both bacteria and their viruses.

For Ma and Bab

ACKNOWLEDGMENTS

For all my wonderful years as a graduate student in the Marraffini lab, I would like to first acknowledge and express my warm admiration for my doctoral advisor, Dr. Luciano Marraffini. Luciano has fostered an inclusive and vibrant community in the lab, and I feel privileged to have been a part of it. As a mentor, Luciano is caring and committed and provided me with the freedom and support to take on explorative projects. He is a thoughtful and brilliant scientist with the ability to simplify complex concepts into digestible ideas, which has helped me tremendously as a student. I have learnt a great deal from Luciano during my time in the lab and am eager to take these lessons into the next steps of my scientific career, wherever they may ultimately lead.

I am sincerely grateful to the members of my thesis committee, Dr. Jeremy Rock and Dr. Paul Bieniasz, for their advice, support and flexibility through the years. I have really benefitted from their feedback during my committee meetings and from their support outside meetings. I am also very grateful to Dr. Graham Hatfull (University of Pittsburgh) for serving as the external examiner for my thesis defense. It is an honor to have a pioneer of contemporary bacteriophage research evaluate my thesis work.

I have made many friends during my time in the Marraffini lab, surrounded by wonderful colleagues through the years. The members of the Marraffini lab, through

all its iterations, have maintained the friendly, fun, and welcoming atmosphere that first drew me to the lab. I would like to acknowledge past and current members of the lab: Dr. Josh Modell, Dr. Wenyan Jiang, Dr. Robert Heler, Dr. Greg Goldberg, Dr. Jon McGinn, Dr. Philip Nussenzweig, Dr. Jakob Rostøl, Dr. Nora Pyenson, Dr. Alex Meeske, Dr. Pascal Maguin, Dr. Andrew Varble, Dr. Charlie Mo, Dr. Claire Kenney, Dr. Dalton Banh, Dr. Naama Aviram, Amanda Shilton, Hyejin Kim, Dr. Raphael Hofmann, Dr. Maj Brodmann, Dr. Paige Arnold, Rahul Bhosle, Phone Pyae-Ko, Albina Kozlova, Ashley Thornal, Jacob Mathai, Christian Baca, Cam Roberts, Gianna Stella, Clare Cahir, Alice Cassel, Calvin Herman, Nia Lyn, Dr. Kailey Slavik and Adriana Mejia, among others, for being amazing colleagues and friends throughout the years. Your support, advice, feedback, and laughs have been an integral presence in my life, both inside and outside the lab.

Of those I mentioned above, I must further acknowledge some for their outsize contributions to my scientific and personal success. First and foremost, I am greatly indebted to Dr. Josh Modell, who was my mentor when I first started in the lab and with whom I collaborated on my first first author scientific publication. I have learned so much from Josh, including many of my better scientific habits, and greatly enjoy his company as a friend. Working with him has been a huge honor. Second, I must acknowledge Christian Baca, a bold and talented graduate student and whom I have had the privilege to mentor in the second half of my PhD. Christian has been an excellent collaborator for my second project, and his insights with protein purification,

biochemistry and bioinformatics have enriched me as a scientist. Third, I would like to thank Adriana Mejia, another talented graduate student who I started mentoring over a year ago and am amazed by how quickly she learns new concepts and techniques. I eagerly await the discoveries Adriana will make in the Marraffini lab. I am also indebted to Dr. Pascal Maguin, Dr. Andrew Varble and Dr. Dalton Banh, for serving as a sounding board for ideas and always providing excellent suggestions and feedback for my research. I have learnt a great deal scientifically from all three of them. On a personal note, many of the colleagues I named earlier have become some of my closest friends outside the lab. Pascal, Andrew, Jakob, Josh, Phillip, Jon, Dalton, Nora, Alex, Christian, Cam, Albina, Hyejin and Claire are some of the most fun and warm people I know and have enjoyed many wonderful moments with them in the lab and outside, around New York City, in our homes and with our family and friends. I am also extremely grateful for all the friends I have made at Rockefeller, including Dr. Sam Khodursky, Dr. Chris Rouya, Dr. Ed Aguilar, Joe Levin, Dr. Chris Cowley, Dr. Zack Mirman, Dr. Peter Mussells Pires, Dr. Dani Keahi, Dr. Lindsey Cantin, Sandra Nakandakari-Higa, Florian Hollunder and Dr. Olivia Goldman, to only name a few. I will forever cherish their friendship.

Rockefeller has been an amazing environment to be a graduate student and I am very grateful to the Dean's Office for their support and flexibility through all these years. I would like to thank Dr. Tim Stearns, Dr. Emily Harms, Marta Delgado, Kristen Cullen, Tasnia Islam, Andrea Morris, Mercedes Harris, and past members of the

Dean's Office, including Cristian Rosario, Stephanie Fernandez, and Dr. Sidney Strickland, for their tireless effort into supporting the student community, accommodating our requests, and ensuring the program runs smoothly. I would also like to thank my collaborators in Dr. Sean Brady's laboratory at Rockefeller, including Dr. Ján Burian, Dr. Vincent Libis and Dr. James Peek, as well as my collaborators Dr. Phoebe Rice, Ying Pigli and Alexis Thomas at the University of Chicago. I would also like to acknowledge staff members at the Rockefeller Genomics Resource Center, particularly Dr. Connie Zhao, Bin Zhang and Christine Lai, and at the Rockefeller Proteomics Resource Center, specifically Dr. Søren Heissel who graciously helped me with challenging mass spectrometry experiments.

On a personal note, I would like to thank my family back home in Bangladesh and in the United States for their unconditional love and support. I am particularly grateful to my parents, Dr. Tasnim Azim and Dr. Ijaz Hossain, for instilling in me a passion for science and for their unwavering love and support during my PhD. They have endured the hard times alongside me, the moments of depression and stress that others do not witness, and I am grateful for their presence in my life. My thesis is first and foremost a dedication to them. Their contributions to society in Bangladesh through their scientific work and their public service is a daily source of inspiration to me, and I hope I have made them proud. I am also grateful to my family in the United States, Fiaz and Seema Hossain, Kameel Mir, Irshad Karim, Kerry Chrystal, Bill and Victoria Karim, and Tariq Ali, for all their support since I moved to the United States in

2010. I am also grateful to my cousins, in Bangladesh and across the world, and whom I all hold dearly in my heart. I would also like to take the opportunity to commemorate my second cousin, Poran, who tragically passed during my first year of graduate school and who we miss dearly every day. Finally, I would like to dedicate this thesis to my grandparents, who instilled in me both a passion for learning and a care for others, and my uncle Bashirul Haq who was a brilliant architect with an unmatched joy for living and who continues to inspire both my aesthetic choices and my drive.

The last three years of my life have been greatly enriched by my wonderful partner, Miriam Gabriel, who is the warmest person I know and is a joy to be with every day. She is unique, creative, free-spirited, funny and a multi-talented artist. She has been my rock during the most trying times. I am also grateful to her parents Janet Heroux and Dr. Abram Gabriel (he too is a molecular biologist), her siblings Ian and Simon Gabriel, and her friends for welcoming me into their lives. Lastly, I would like to thank my friends – friends in Bangladesh, friends from Bangladesh in New York, friends from New York, and friends from college – who are too many to name but have made my life exciting and enriching through all my years of graduate school.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2-aminoA……..2-aminoadenine

abi……abortive infection

AdfA……..anti-DarT factor A

a-GT……alpha-glucosyltransferase protein

a-gt……..alpha-glucosyltransferase gene

Acb……anti-CBASS

Acr……anti-CRISPR

ADP……...adenosine diphosphate

AP……...apurinic/apyrimidinic

Apyc……anti-Pycsar

Ara…….....L-arabinose

ATP……...adenosine triphosphate

Avs……....antiviral STAND

b-GT……..beta-glucosyltransferase protein

b-gt……beta-glucosyltransferase gene

ba-GT……beta-alpha glucosyltransferase protein

ba-gt……...beta-alpha glucosyltransferase gene

B. subtilis……*Bacillus subtilis*

BER……base excision repair

bp……base pair

BREX……bacteriophage exclusion

Brig1…….bacteriophage replication inhibition DNA glycosylase 1

Cas……CRISPR-associated

Cascade……..…….CRISPR-associated complex for antiviral defense

CBASS……...….cyclic oligonucleotide-based anti-phage signaling system

CFU…….colony forming unit

chi.…….crossover hotspot instigator

CRISPR…….……clustered regularly interspaced short palindromic repeat

crRNA……..…….CRISPR RNA

Da…….....dalton

dCas9…….....dead Cas9

DISARM……..defense island system associated with restriction-modification

dATP…….……...deoxyadenosine triphosphate

dCTP…….…….....deoxycytidine triphosphate

dGTP…….……...deoxyguanosine triphosphate

DNA……deoxyribonucleic acid

dNTP…….……...deoxynucleotide triphosphate

DSB…….……DNA double-strand break

dsDNA……double-stranded DNA

dsRNA……double-stranded RNA

DSR…….……..defense-associated sirtuin

dTTP…….……....deoxythymidine triphosphate

dUMP…….……..deoxyuridine monophosphate

dUTP……deoxyuridine triphosphate

E. coli……*Escherichia coli*

Endo IV……Endonuclease IV

Fe-S……iron-sulfur

Glc……glucosyl

gp……gene product

hmC……5-hydroxymethylcytosine

hmdC……5-hydroxymethyldeoxycytidine

HR……homologous recombination

hSMUG1……human single-strand selective monofunctional uracil DNA glycosylase 1

IPI*……Internal Protein I*

IPTG……isopropyl β-D-1-thiogalactopyranoside

kb……kilobase

L……DNA size ladder

LB……lysogeny broth

Lig……ligase

mC……5-methylcytosine

mins……minutes

MMEJ……microhomology-mediated end joining

MOI……multiplicity of infection

MS……mass spectrometry

msDNA……multi-copy single-stranded DNA

MW……molecular weight

NAD……nicotinamide adenine dinucleotide

NAD^+……oxidized form of NAD

NaOH……sodium hydroxide

NER……nucleotide excision repair

Nfo……Endonuclease IV

NGS……next-generation sequencing

NHEJ……non-homologous end joining

Nhi……nuclease-helicase immunity

NLR……nucleotide-binding oligomerization domain-like receptor

NT……non-targeting

nt……nucleotide

Nth……Endonuclease III

oligo……oligonucleotide

OD_{600}……optical density at wavelength 600 nanometers

p……plasmid or cosmid

PAGE……polyacrylamide gel electrophoresis

pAgo……prokaryotic argonaute

PAM……protospacer adjacent motif

PAMP……pathogen-associated molecular pattern

PARIS……phage anti-restriction-induced system

PCR……polymerase chain reaction

PFU……plaque forming unit

Pgl……..phage growth limitation

phage……..bacteriophage

Pol……..polymerase

Pycsar……..……pyrimidine cyclase system for anti-phage resistance

qPCR……..quantitative PCR

RM……..……..restriction–modification

RNA…….……ribonucleic acid

S. pyogenes……..…*Streptococcus pyogenes*

S. tokodaii……..……...…*Sulfolobus tokodaii*

SD……..standard deviation

SDS……..sodium dodecyl sulfate

SEM……standard error of the mean

Sie…….……superinfection exclusion

SIR2……..sirtuin

Sm……..hSMUG1

spc……..…….....spacer

SSB……single-stranded binding protein

ssDNA…….……single-stranded DNA

STAND……...signal transduction ATPases with numerous domains

TA……toxin-antitoxin

Tad……Thoeris anti-defense

TIR……Toll/interleukin-1 receptor

ThsA……...Thoeris A

tracrRNA……trans-activating CRISPR RNA

tRNA……transfer RNA

U……uracil

UV……..ultraviolet

vir……..virulent mutant

WT……..wild-type

wt……wild-type

XthA……Exonuclease III

CHAPTER 1. INTRODUCTION

1.1 Genetic conflicts between bacteria and their viruses

Genetic conflicts provide a dynamic avenue for molecular innovation, through the evolution of novel mechanisms of attack and defense. Evolutionary biologist Leigh Van Valen's Red Queen Hypothesis states that biological conflicts are locked in dynamic equilibrium through constant ongoing evolutionary adaptations by both sparring biological entities (Liow et al., 2011). In nature, this is best typified by the evolutionary conflicts between bacteria and their viruses. Prokaryotic host-virus genetic conflicts have been raging since ancient evolutionary time and have informed much of our current understanding of molecular biology. While many of these studies were performed decades ago to base the foundations of current biological thinking, the study of these conflicts is undergoing a renaissance with the recent discovery of numerous, diverse bacterial immune systems, including CRISPR-Cas (Bernheim and Sorek, 2020; Georjon and Bernheim, 2023; Rostol and Marraffini, 2019a). Indeed, what is increasingly apparent is that genetic conflicts between bacteria and their viruses drive unexpected molecular inventions and present exciting avenues for the discovery of new molecular biology and new biotechnology tools, such as Cas9 used for gene editing (Cong et al., 2013).

The conflicts between bacteria and their viruses have resulted in a diverse array of bacterial immune systems (Georjon and Bernheim, 2023). Since several of

these systems sense and target the DNA of invading viruses, I hypothesized that DNA repair proteins may play an important role in these molecular skirmishes. In this thesis, I will investigate how both bacteria and their viruses co-opt DNA repair proteins, which canonically function in genome repair and integrity, to execute non-canonical immune functions in prokaryotic host-virus genetic conflicts.

1.2 Bacteriophages

Bacterial viruses are called bacteriophages or phages for short. They were discovered in the early twentieth century, independently by Félix d'Hérelle and Frederick W. Twort, and named for their ability to "devour" bacteria (d'Herelle, 1917; Twort, 1915). They represent the most abundant biological entities on Earth, estimated at about 10^{31} phage particles on the planet's surface (Brüssow and Hendrix, 2002). Given their ubiquity, phages unexpectedly possess vast morphological diversity (Ackermann, 2007; Dion et al., 2020). Fundamentally, they consist of a core of nucleic acids surrounded by a proteinaceous shell called the capsid. In many cases, the phage virion will have a protein tube called a tail attached to the capsid. The tail can be of variable length, may have shorter tail fibers and facilitates adsorption of the viral particle to the outer surface of its bacterial host. The nucleic acid of the virus can be DNA or RNA, either single-stranded or double-stranded (Ackermann, 2007; Dion et al., 2020). Most phages that have been isolated so far belong to the order *Caudovirales* of the class *Caudoviricetes* (Ackermann, 2007). These are a class of tailed bacteriophages with icosahedral capsids and

double-stranded DNA (dsDNA) genomes, ranging from 18-50 kilobases (kb) in size (Dion et al., 2020) (**Fig. 1.1**). Most phages within the order *Caudovirales* belong to the family *Siphoviridae*, which include the well-known *Escherichia coli* phages λ and T5. Siphophages have nonenveloped capsids with long noncontractile tails and linear dsDNA genomes (Dion et al., 2020). Other well-represented families within the order are *Myoviridae* and *Podoviridae*, which also possess nonenveloped capsids and linear dsDNA genomes (Ackermann, 2007; Dion et al., 2020; Ofir and Sorek, 2018). Myophages have contractile tails and are best exemplified by the T-even phages that infect *E. coli*, of which phage T4 is the best described. Podoviruses, in contrast, have very short, noncontractile tails and include the highly virulent coliphages T3 and T7.

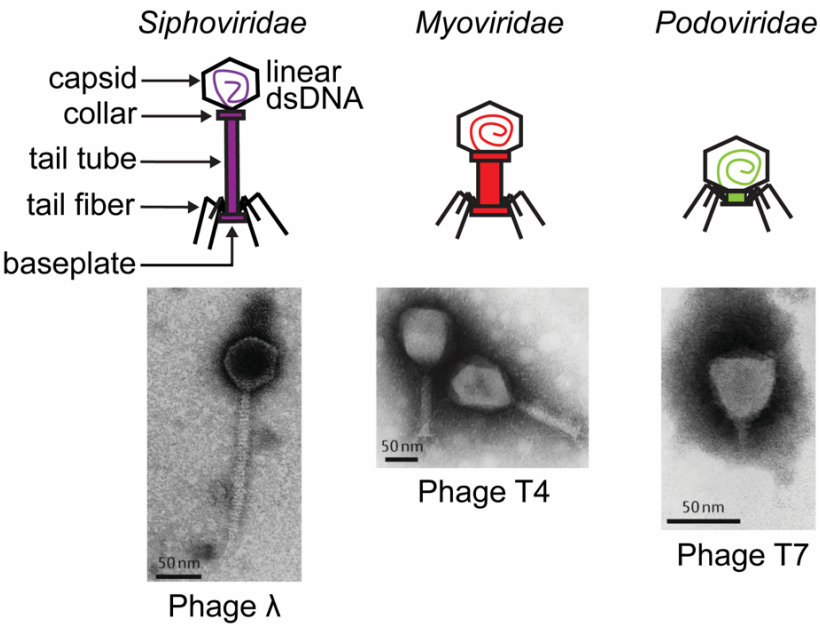


Figure 1.1 The three major families of bacteriophages in the order *Caudovirales*

Morphologies and transmission electron micrographs of representative coliphages in the families *Siphoviridae*, *Myoviridae* and *Podoviridae*, within the order *Caudovirales*. Phages within this order have icosahedral capsids that encapsidate linear dsDNA genomes. Transmission electron micrographs of phages λ, T4 and T7 are from Dion et al. (2020) and reproduced with permission from Springer Nature.

1.3 The phage lifecycle: lysis or lysogeny

Like other viruses, phages are obligate parasites: they require a bacterial cell to carry out their lifecycle. Once a phage has adsorbed to its target bacterium, it injects its genetic material into the cell's cytosol. Inside the cell, the phage lifecycle can be lytic or lysogenic (Herskowitz and Hagen, 1980) (**Fig. 1.2**). Lytic phages, including coliphages T3, T4, T5 and T7, are only able to carry out the lytic cycle, whereby infected cells lyse to release viral progeny (**Fig. 1.2-A**). In the lytic cycle, the injected linear dsDNA of the phage immediately circularizes to form a closed circle. This avoids degradation of the linear dsDNA ends by host nucleases. The closed dsDNA circle replicates rapidly to produce multiple monomeric copies via theta replication. The replicative mode ultimately switches to rolling-circle replication, where monomeric circles are replicated to produce concatemeric phage genomes which are concurrently packaged into viral capsids to produce mature virions. Lysis of the infected cell, driven by phage-produced lysin proteins, releases the encapsidated phage genomes, enabling the next round of infection (Ofir and Sorek, 2018). In our laboratory experiments, a single round of lysis typically results in 50-200 infectious virions per infected cell for the phages mentioned above.

Temperate phages, such as phage λ, can undergo either lysis or lysogeny (Herskowitz and Hagen, 1980), and the lysis-lysogeny decision is dictated by a molecular switch (Oppenheim et al., 2005). During lysogeny, the phage foregoes the lytic cycle and instead integrates its genetic material into the chromosome of its

bacterial host. The integrated viral genome is described as a prophage and remains as such, in a dormant state and undergoing bacterial cell division, until and unless the prophage receives an environmental or molecular cue that triggers its excision out of the chromosome. Following excision, the phage immediately circularizes and proceeds with the viral lytic cycle (**Fig. 1.2-B**).

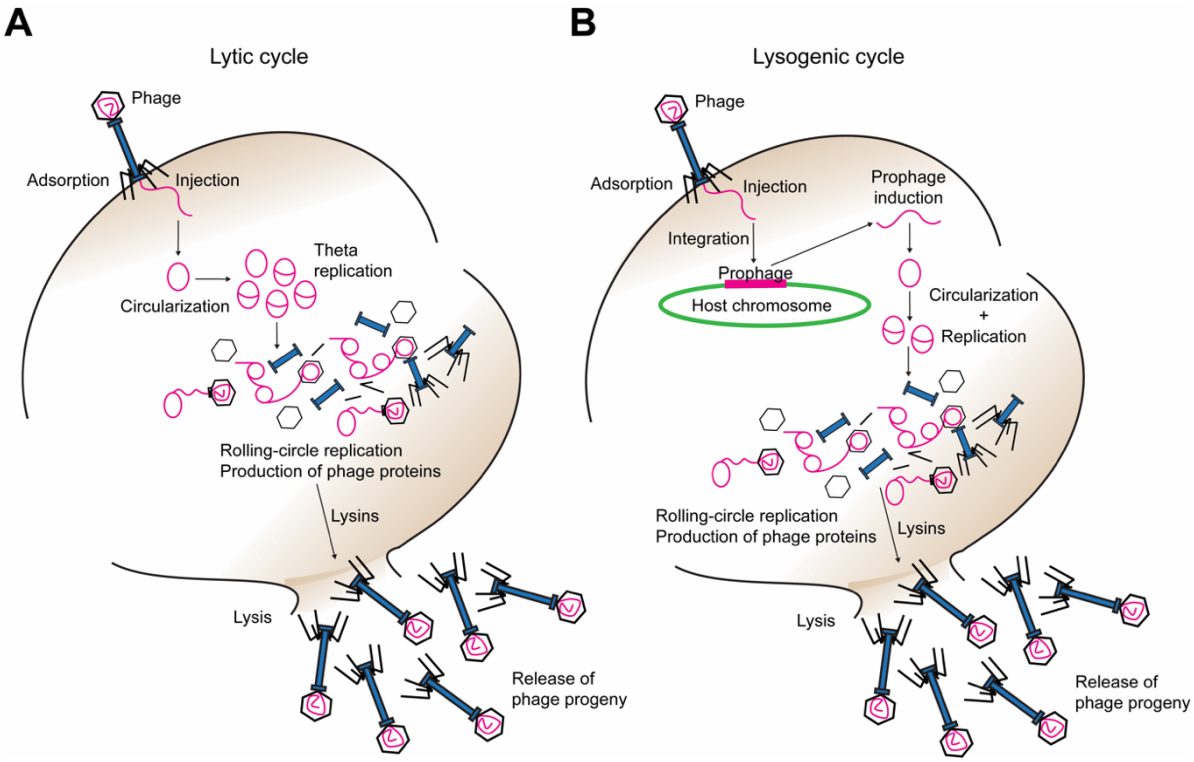


Figure 1.2 The phage lifecycle: lysis and lysogeny

(**A**) In the lytic cycle, the phage genome immediately circularizes upon injection and undergoes theta replication to produce monomeric circular genomes. Monomeric circles switch to rolling-circle replication to produce linear concatemeric phage genomes that can be packaged into capsids. Production of lysins and other phage proteins during late infection cause cell lysis, releasing the packaged phage genomes, which can now go on to infect neighboring cells. (**B**) In the lysogenic cycle, the phage forgoes lysis, and instead integrates its genome into the host chromosome. The integrated phage genome is called a prophage, and the prophage-encoding bacterium is called a lysogen. A molecular or environmental cue may lead to the prophage excising out of the bacterial chromosome in a process called prophage induction. The excised phage genome circularizes and undergoes the lytic cell to spread phage progeny through cell lysis.

1.4 Antiviral immune systems in bacteria: discovery and diversity

Phages are the most abundant biological entities in Earth's biosphere and are estimated to outnumber bacteria 10:1 in the ocean (Parikka et al., 2017). Indeed, a large fraction of daily bacterial deaths, 20-40%, are believed to be driven by phage infections (Suttle, 2007). To survive this daily onslaught, bacteria have evolved diverse strategies to counter phages. Bacterial anti-phage strategies include preventing phage adsorption or injection into the cell, targeting the phage's nucleic acids, blocking viral replication or transcription, and sensing phage infection to drive abortive infection or induce cellular dormancy, both strategies that prevent or minimize the release of infectious viral progeny (Rostol and Marraffini, 2019a) (**Fig. 1.3**). Until recently, much of our knowledge of bacterial anti-phage systems was limited largely to restriction-modification (RM) and CRISPR-Cas systems, given their use as tools for molecular cloning and gene editing. Recent bioinformatics approaches, however, have led to an explosion in the discovery of new, diverse anti-phage systems in bacteria, illustrating that the prokaryotic world possesses a rich immune repertoire (Doron et al., 2018; Fillol-Salom et al., 2022; Gao et al., 2020; Millman et al., 2022; Rousset et al., 2022).

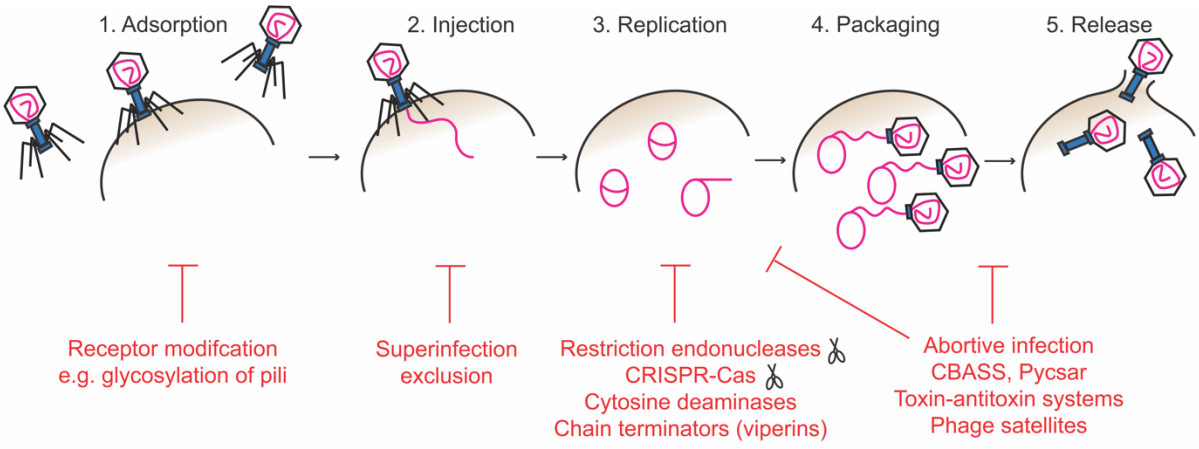


Figure 1.3 Bacterial anti-phage defense strategies target different stages of the phage lifecycle

Bacteria use various strategies to target different stages of the viral lifecycle. Receptor modification and superinfection exclusion prevent adsorption or injection of the phage. A large diversity of bacterial immune systems prevent viral replication, some of which cleave phage nucleic acids such as CRISPR-Cas and restriction endonucleases, while others deplete nucleotides (e.g. cytosine deaminases) or function as replication chain terminators (e.g. prokaryotic viperins). Finally, many abortive infection systems are activated late during viral infection and cause cell suicide or cell arrest, thereby prevent viral packaging and the release of infectious viral progeny.

Anti-phage defense systems tend to colocalize in bacterial genomes, in hotspots called “defense islands” (Makarova et al., 2013; Makarova et al., 2011). Bioinformatics approaches have exploited this tendency by using a “guilt-by-association” strategy to identify undiscovered prokaryotic immune systems within defense islands in published bacterial genomes available in online databases (Doron et al., 2018; Gao et al., 2020; Johnson et al., 2022; Millman et al., 2022). The putative anti-phage systems, which can be either single-gene systems or multi-gene operons, are then validated experimentally, by cloning in the system into a model

organism, such as *E. coli* or *Bacillus subtilis*, and challenging the bacterium with a lytic phage (**Fig. 1.4**). A similar approach has focused on accessory genes carried in prophage genomes and pathogenicity islands (Fillol-Salom et al., 2022; Rousset et al., 2022), since these often encode anti-phage defense systems including superinfection exclusion (Sie) systems, such as the RexAB system in phage λ (Lopatina et al., 2020), which prevent infection by the same phage or competing phages. Following these bioinformatic pipelines, an amazing diversity of bacterial immune systems have been discovered since 2018, with emerging parallels between prokaryotic and eukaryotic immunity (Georjon and Bernheim, 2023).

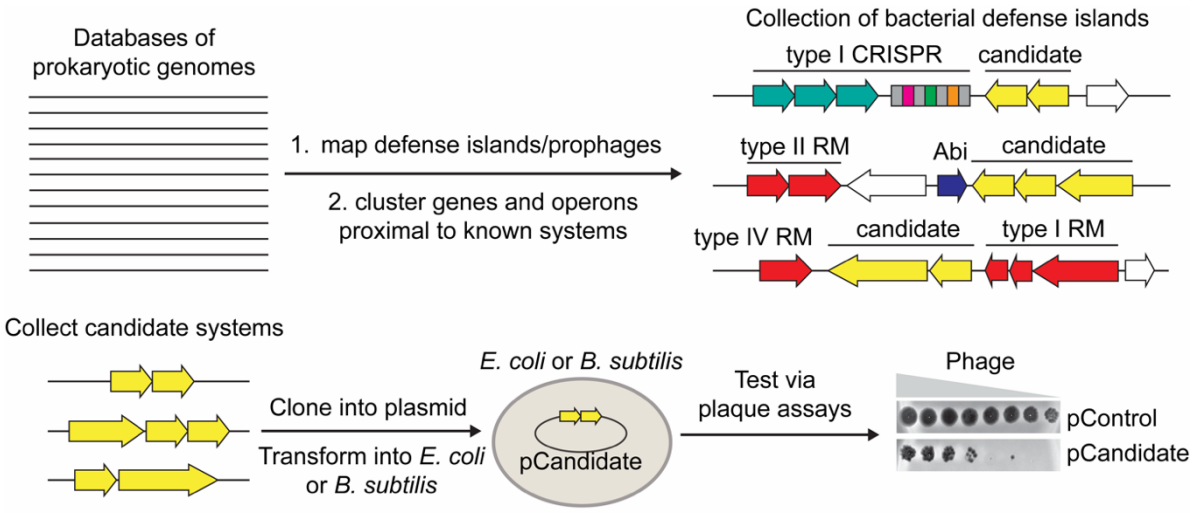


Figure 1.4 Bioinformatics approaches to identify and test novel candidate anti-phage defense systems

Bioinformatics approaches to identify novel candidate bacterial immune systems scan through online databases of published bacterial genomes and map defense islands and prophages within these genomes. They then cluster unknown genes and operons proximal to known defense systems or ones that frequently occur in prophage genomes. The candidate defense systems are cloned into a model organism, such as *E. coli* or *B. subtilis*, and tested for immunity via a plaque assay. Figure primarily highlights the approach of Doron et al. (2018).

More recently, Vassallo et al. performed a functional selection of the *E. coli* pangenome to discover anti-phage defense systems across all isolated *E. coli* strains (Vassallo et al., 2022). Their work uncovered 21 new anti-phage defense systems that had gone undetected from bioinformatics approaches that have prioritized guilt-by-association with defense islands and prophages. Therefore, while many anti-phage systems have already been discovered and characterized, this study suggests that there remain several unknown bacterial immune systems that still await our discovery.

At the molecular level, anti-phage systems are incredibly varied and comprise assorted protein domains with diverse activity and function (Bernheim and Sorek, 2020; Georjon and Bernheim, 2023). The domain, protein or complex that executes the antiviral action is called the effector. Anti-phage effectors may either directly attack the phage, for example by cleaving its nucleic acids, or induce host cell dormancy or abortive infection, for example through a nonspecific RNase that indiscriminately degrades host and viral transcripts (Rostol and Marraffini, 2019b), to slow host cell activity and/or promote premature cell death to minimize the release of phage progeny (Lopatina et al., 2020). In many cases, the anti-phage effector is accompanied by a sensor, which is a domain, protein or complex that detects phage infection, often through direct recognition of a distinct molecular feature of the virus called a pathogen-associated molecular pattern (PAMP) (Janeway, 1989). Many anti-phage systems detect or target viral nucleic acids (Georjon and Bernheim, 2023),

while others detect phage structural proteins, such as capsids or tails (Gao et al., 2022; Garb et al., 2022; Zhang et al., 2022), phage effector proteins that promote virulence (Rousset et al., 2022) or the disruption/dysregulation of host metabolites and processes during phage infection (LeRoux and Laub, 2022) (**Fig. 1.5**). The different modes of bacterial anti-phage immunity are described in the sections below.

1.5 Preventing phage adsorption and injection

Since the initiation of infection requires the phage to insert its genetic material into the host cell, many bacteria employ strategies aimed at preventing viral adsorption or injection (Rostol and Marraffini, 2019a). A common strategy for preventing phage adsorption is to modify surface proteins, since phages require surface receptors to adsorb to target cells. An example is the glycosylation of type IV pili in *Pseudomonas aeruginosa* to prevent adsorption of pilus-dependent phages (Harvey et al., 2018). Sie mechanisms can also prevent phage injection. For example, the mycobacterial phage Fruitloop expresses the protein gp52 which inactivates the host's cell wall synthesis protein Wag31 (Ko and Hatfull, 2018). An unrelated group of phages, that use Wag31 for DNA injection, are now excluded from infecting the same host. Finally, a common and simple strategy to prevent adsorption or injection is for the bacteria to mutate the surface receptor that the phage uses. This is observed during phage λ infection in *E. coli*, where mutations in the receptor LamB affect phage adsorption (Clement et al., 1983).

1.6 Abortive infection mechanisms

Once the phage enters the cell, it can begin its lifecycle. For lytic phages, this involves replication of the phage's genetic material and the transcription and translation of phage proteins (**Fig. 1.2-A**). These processes disrupt cellular metabolite levels and normal host activities, for example through shutoff of host transcription (Drivdahl and Kutter, 1990) and depletion of key metabolites such as ATP and NAD^+. In many instances, phages express effector proteins that promote virulence, for example proteins that inhibit host recombinases (Murphy, 1991; Williams and Radding, 1981) and restriction enzymes (Walkinshaw et al., 2002). Viral proteins, nucleic acids, and infection-associated processes all provide avenues for the detection of phage infection by bacterial anti-phage defense systems. In abortive infection systems, detection of phage infection activates an anti-phage effector, whose activity promotes the arrest or death of the infected cell, thereby preventing the release and spread of phage progeny (Lopatina et al., 2020). Recent discoveries, driven by bioinformatics, have shed light on the vast diversity of abortive infection systems in bacteria (Georjon and Bernheim, 2023) (**Fig. 1.5**).

Among the most well-represented abortive infection systems in bacterial genomes are CBASS (cyclic oligonucleotide-based anti-phage signaling system) and retrons (Georjon and Bernheim, 2023) (**Fig. 1.6**). In CBASS systems, a cyclase senses phage infection, in some instances through detection of a structured viral RNA produced by the target phage (Banh et al., 2023), to generate cyclic di-

/trinucleotides that activate a downstream immune effector (Cohen et al., 2019; Millman et al., 2020b). The cyclase bears resemblance to the eukaryotic innate immune sensor, cGAS (cyclic GMP-AMP synthase), which produces cyclic GMP-AMP to activate STING (stimulator of interferon genes) in response to dsDNA or dsRNA in the cytosol of eukaryotic cells (Ablasser et al., 2013; Slavik et al., 2021; Sun et al., 2013). Activation of STING mounts an interferon response in the cell (Ishikawa et al., 2009).

Akin to CBASS, other signaling-based anti-phage systems include Pycsar (pyrimidine cyclase system for anti-phage resistance) and Thoeris, which produce cyclic pyrimidine and cyclic ADP-ribose molecules, respectively, to activate downstream effectors (Ofir et al., 2021; Tal et al., 2021). While the phage cues for most CBASS, Pycsar and Thoeris systems are unknown, many anti-phage abortive infection systems have been shown to directly sense phage structural proteins or effectors. For example, Avs, antiviral proteins of the STAND superfamily, have nucleotide-binding oligomerization domain-like receptor (NLR) domains (Gao et al., 2022). NLR domains act as pattern recognition receptors in animal inflammasomes and plant resistosomes (Jones et al., 2016). Avs homologs in bacteria are anti-phage abortive infection systems that detect phage terminase or portal proteins to activate varied effector domains (Gao et al., 2022). Avs are single-gene systems, as is the CapRel toxin-antitoxin system that binds a phage capsid protein to relieve toxin repression by its antitoxin domain (Zhang et al., 2022). The liberated toxin domain

pyrophosphorylates tRNAs to inhibit translation, leading to cell dormancy or death. Similarly, defense-associated sirtuins (DSRs) detect the phage tail tube protein to trigger abortive infection, through an effector that depletes the metabolite NAD^+ in the host cytosol (Garb et al., 2022). While these systems recognize PAMPs, PARIS (phage anti-restriction-induced system), an abortive infection system found in multiple prophages, is activated upon sensing the viral anti-restriction protein Ocr, i.e., a phage effector that inhibits RM and BREX (bacteriophage exclusion) systems (Rousset et al., 2022). PARIS is described as an “anti-anti-restriction” system and constitutes an example of effector-triggered immunity, whereby a viral effector stimulates an immune response in the infected host (Remick et al., 2023). PARIS beautifully exemplifies the myriad layers of evolution that define prokaryotic host-virus conflicts, since the system represents a bacterial counter strategy to its target phage’s inhibition of critical host antiviral systems.

As mentioned earlier, retrons are among the more common abortive infection systems encoded in prokaryotic genomes (Georjon and Bernheim, 2023; Tesson et al., 2022). The anti-phage mechanism of retrons constitute an example of “guard theory”, since retrons monitor the integrity of host cellular processes, which are commonly dysregulated during viral infection. Many retrons “guard” the bacterial recombinase-nuclease RecBCD. In retron systems, a multi-copy single-stranded DNA (msDNA), or a chimeric DNA/RNA derived from the msDNA, acts as an antitoxin that represses the retron effector (Bobonis et al., 2022; Millman et al.,

2020a). Inhibition of RecBCD by a phage protein such as λ Gam or methylation of the msDNA by a phage-encoded methyltransferase disrupts the retron msDNA and relieve inhibition of the retron's effector subunit, driving cell arrest. Other bacterial toxin-antitoxin systems similarly monitor infection-associated processes and trigger abortive infection effectors (toxins) in response to phage infection. For example, the ToxIN and RnlAB toxin-antitoxin systems detect inhibition of transcription and translation, respectively, to activate RNase toxins that mediate indiscriminate RNA cleavage (Guegler and Laub, 2021; Koga et al., 2011). Similarly, dCTP deaminases and dGTPases in bacterial defense islands, which deplete deoxycytidine and deoxyguanosine nucleotides, respectively, are triggered by host transcription shutoff during phage infection, although the exact mechanisms of activation are currently unclear (Hsueh et al., 2022; Tal et al., 2022).

The sensors in bacterial abortive infection systems activate a wide range of immune effectors. These effectors are diverse in domain and activity. They target nucleic acids, through DNA/RNA endonucleases and helicases, target proteins, through proteases, deplete metabolites, through deaminases, ATPases, dNTPases and NADases, and disrupt membranes, via transmembrane proteins, phospholipases, and pore-forming proteins (Georjon and Bernheim, 2023). While abortive infection systems are diverse and widespread, a more common mechanism of anti-phage immunity entails directly targeting the DNA of invading phages (Georjon and Bernheim, 2023) (**Fig. 1.6**).

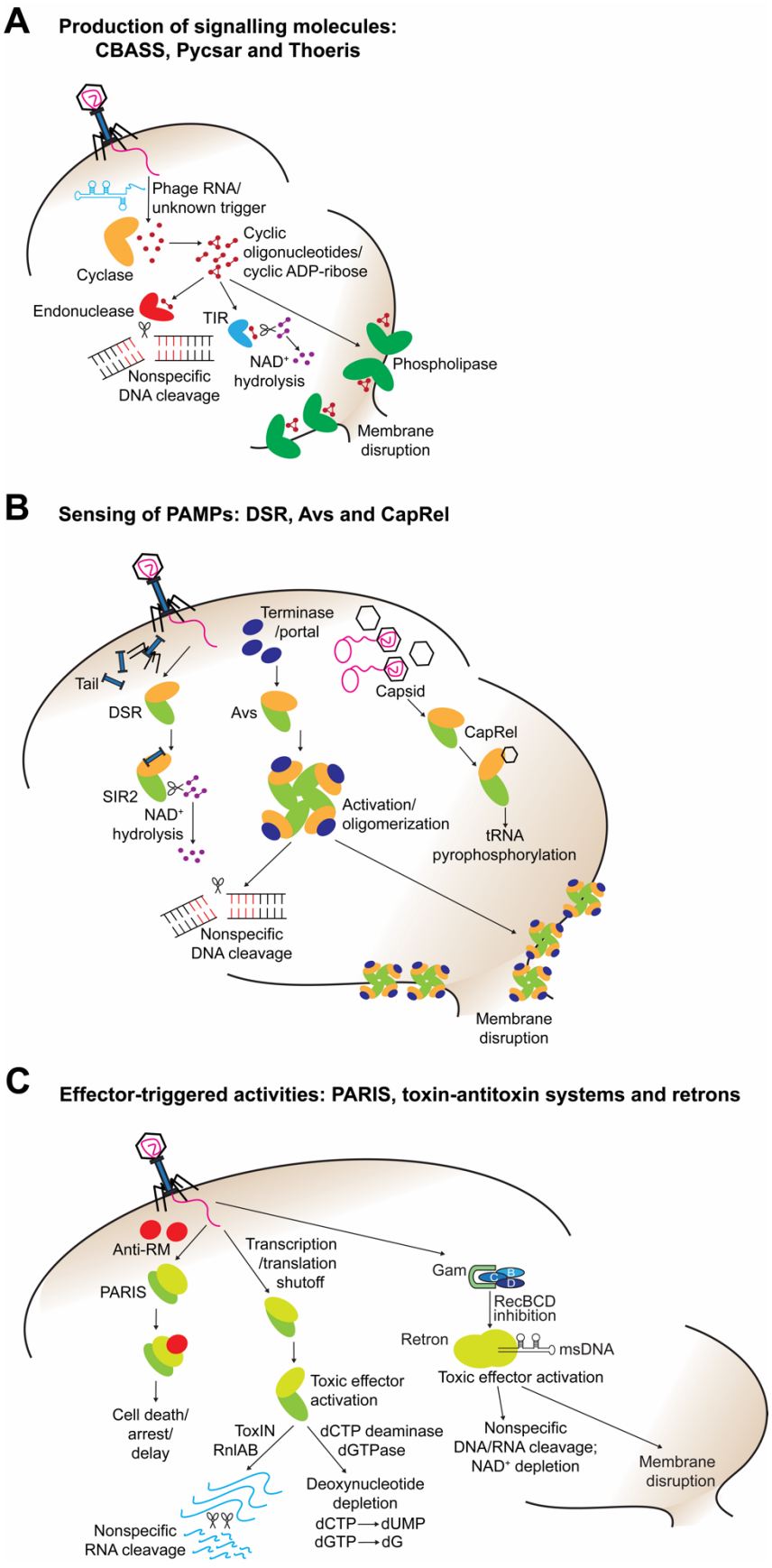

A
Production of signalling molecules:
CBASS, Pycsar and Thoeris
Phage RNA/
unknown trigger
Cyclase
Cyclic
oligonucleotides/
cyclic ADP-ribose
Endonuclease
TIR
NAD+
hydrolysis
Nonspecific
DNA cleavage
Phospholipase
Membrane
disruption
B
Sensing of PAMPs: DSR, Avs and CapRel
Terminase
/portal
Tail
DSR
Avs
Capsid
CapRel
SIR2
NAD+
hydrolysis
Activation/
oligomerization
tRNA
pyrophosphorylation
Nonspecific
DNA cleavage
Membrane
disruption
C
Effector-triggered activities: PARIS, toxin-antitoxin systems and retrons
Anti-RM
PARIS
Transcription
/translation
shutoff
Gam
RecBCD
inhibition
Retron
msDNA
Toxic effector activation
Cell death/
arrest/
delay
Toxic effector
activation
ToxIN
RnlAB
dCTP deaminase
dGTPase
Deoxynucleotide
depletion
dCTP⟶dUMP
dGTP⟶dG
Nonspecific
RNA cleavage
Nonspecific
DNA/RNA cleavage;
NAD+ depletion
Membrane
disruption

Figure 1.5 Abortive infection systems in bacteria

(**A**) Immune systems such as CBASS, Pycsar and Thoeris encode cyclases that produce cyclic oligonucleotides or cyclic ADP-ribose upon detection of phage infection. These signaling molecules activate a variety of downstream toxic effectors, including endonucleases, NAD^+-degrading TIR proteins and membrane disrupting phospholipases. (**B**) Many abortive infection systems detect phage PAMPs, including tail, terminase, portal and capsid proteins. PAMP binding leads to the activation of toxic effector domains that hydrolyze NAD^+, cleave nucleic acids, disrupt membranes or pyrophosphorylate tRNAs. (**C**) Many abortive infection pathways are activated by phage-encoded effector activities. The PARIS system induces cell arrest or death upon detecting the anti-RM protein Ocr produced by phage T7. Shutoff of host transcription activates anti-phage dCTP deaminases and dGTPases as well as the ToxIN toxin-antitoxin system which mediates nonspecific RNA cleavage. Shutoff of host translation activates the RnlAB toxin-antitoxin system, which also cleaves RNAs. Retrons are commonly activated by inhibition of the host recombinase-nuclease complex RecBCD and trigger a diversity of downstream toxic effectors. Phage-encoded Gam proteins inhibit RecBCD, thereby activating retron-mediated abortive immunity. Figure modified from Georjon and Bernheim (2023) and reproduced with permission from Springer Nature.

1.7 Bacterial anti-phage systems that target viral DNA

Targeting a virus's nucleic acids is an effective strategy for thwarting its replication. Since most phages that have been isolated or sequenced thus far possess dsDNA genomes, it is unsurprising that the most abundant anti-phage systems directly target viral DNA. Indeed, RM and CRISPR-Cas, both DNA-targeting immune systems, are heavily overrepresented in prokaryotic genomes (Georjon and Bernheim, 2023) (**Fig. 1.6**).

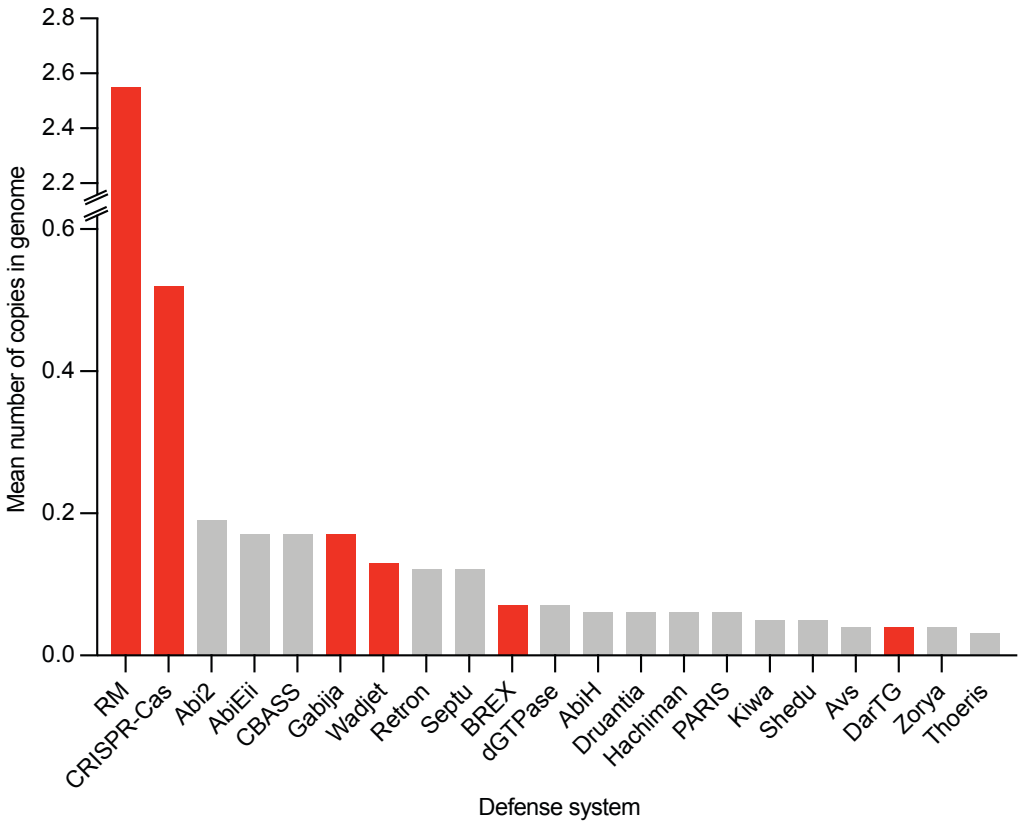


Figure 1.6 Occurrence of common anti-phage defense systems in bacterial genomes

Average number of copies of each indicated anti-phage defense system encoded in all published bacterial genomes. Red bars indicate DNA-targeting anti-phage systems. Figure modified from Georjon and Bernheim (2023) and reproduced with permission from Springer Nature.

RM and CRISPR-Cas systems recognize target sequences within invading foreign DNA, including both phages and plasmids. Target recognition is often dependent on the presence or absence of DNA nucleobase modifications, which provide a mechanism for the system to distinguish between self and non-self DNA (**Fig. 1.7-AB**) (Tock and Dryden, 2005). While RM systems cleave dsDNA upon target recognition (**Fig. 1.8-A**), CRISPR-Cas systems may be more varied in their

effector activity (Nussenzweig and Marraffini, 2020). CRISPR-Cas system rely on guide RNAs to recognize their target sequences (**Fig. 1.7-C**) (Marraffini, 2015). Many Cas effectors cleave ssDNA or dsDNA through nuclease (and helicase) activities, resulting in the direct degradation of phage or plasmid DNA (**Fig. 1.8ACDE**) (Nussenzweig and Marraffini, 2020). However, unlike RM systems, some CRISPR-Cas effectors also mediate abortive infection through varied mechanisms, including indiscriminate degradation of host and viral nucleic acids — via nonspecific RNase and ssDNase effectors (Jiang et al., 2016; Meeske et al., 2019; Rostol and Marraffini, 2019b; Rostol et al., 2021) — and membrane disruption — via transmembrane pore-forming effectors (VanderWal et al., 2023). Similarly, prokaryotic argonautes (pAgos) are DNA-targeting abortive infection systems found in some bacteria, although they are far less common than RM or CRISPR-Cas. pAgos use short fragments of DNA derived from the genomes of invading plasmids and phages as guide DNAs to recognize target sequences and trigger cell arrest, through NAD^+ depletion or membrane disruption (**Fig. 1.7-C**) (Zaremba et al., 2022; Zeng et al., 2022).

Like RM, the anti-phage systems BREX (Goldfarb et al., 2015; Gordeeva et al., 2018) and DISARM (defense island system associated with restriction-modification) (Ofir et al., 2018) rely on the differential methylation of host and foreign DNA to recognize their targets (**Fig. 1.7-A**). Both systems carry DNA methylases that methylate host DNA, thus registering any invading DNA that lacks the appropriate DNA methyl marks as non-self. While these systems encode ATPases or helicases,

among other genes, their exact mechanism of targeting foreign DNA remains unclear. In a similar vein, the DndABCDE- and SspABCD-based anti-phage systems introduce sulfur modifications into DNA sugar-phosphate backbones as a means of self/non-self discrimination (**Fig. 1.7-A**) (Wang et al., 2007; Wang et al., 2021; Xiong et al., 2019; Xiong et al., 2020; Xu et al., 2010). The effector modules of these systems, DndFGH and SspE or SspFGH, respectively, cleave or damage unmodified phage DNA, although further work is necessary to disentangle their precise mechanisms of action. Another example of a DNA-modifying anti-phage system is the toxin-antitoxin system DarTG (LeRoux et al., 2022). While the systems described above exploit DNA modifications for self/non-self discrimination, the DarT toxin is a DNA-modifying anti-phage effector. During infection by a target phage, DarT is liberated from its antitoxin DarG and subsequently ADP-ribosylates thymine bases in DNA as a means of obstructing phage DNA replication (**Fig. 1.8-F**).

A significant downside to DNA modification-based discrimination is that the targeted phage or plasmid can gain or lose DNA modifications through both genetic and epigenetic pathways, providing an easy route for immune evasion (Bickle and Kruger, 1993; Maguin et al., 2022; Samson et al., 2013; Tock and Dryden, 2005). As such, some bacterial immune systems instead exploit the differential DNA topologies that exist in phages and plasmids to specifically target these mobile genetic elements over the host chromosome (**Fig. 1.7-D**). For example, the single enzyme system Nhi (nuclease-helicase immunity) in Staphylococci is activated by a phage-encoded

single-stranded DNA binding protein (SSB) and selectively degrades phage replication intermediates while leaving host DNA undamaged (Bari et al., 2022). The multi-gene Wadjet defense system also exploits differences in DNA topology to defend bacteria against invading plasmids (Deep et al., 2022; Liu et al., 2022). The Wadjet complex specifically recognizes and cleaves closed-circular plasmid DNA, without attacking host chromosomal DNA, which possesses a different topology to plasmid DNA. Three Wadjet subunits form a structural maintenance of chromosome (SMC) complex, which acts as a DNA-binding motor that hydrolyzes ATP and activates a homodimeric endonuclease subunit, with homology to type II DNA topoisomerases, to cleave the plasmid DNA.

The systems described here specifically recognize or target foreign DNA, although their modes of recognition (**Fig. 1.7**) and their mechanisms of generating DNA damage (**Fig. 1.8**) are quite varied. In addition to these, many DNA-targeting proteins function as abortive infection effectors, including in CBASS (Lau et al., 2020; Lowey et al., 2020) and Avs systems (Gao et al., 2022). These systems commonly encode DNA endonuclease domains of the PD-(D/E)XK, Mrr, Cap4 and HNH families (Gao et al., 2022; Lau et al., 2020; Lowey et al., 2020) that mediate indiscriminate cleavage of both host and viral DNA upon activation (**Fig. 1.5**). Overall, bacterial immune systems that sense and target phage DNA are plentiful and diverse. Next, I will review the two most abundant bacterial defense systems, RM and CRISPR, both of which target foreign DNA.

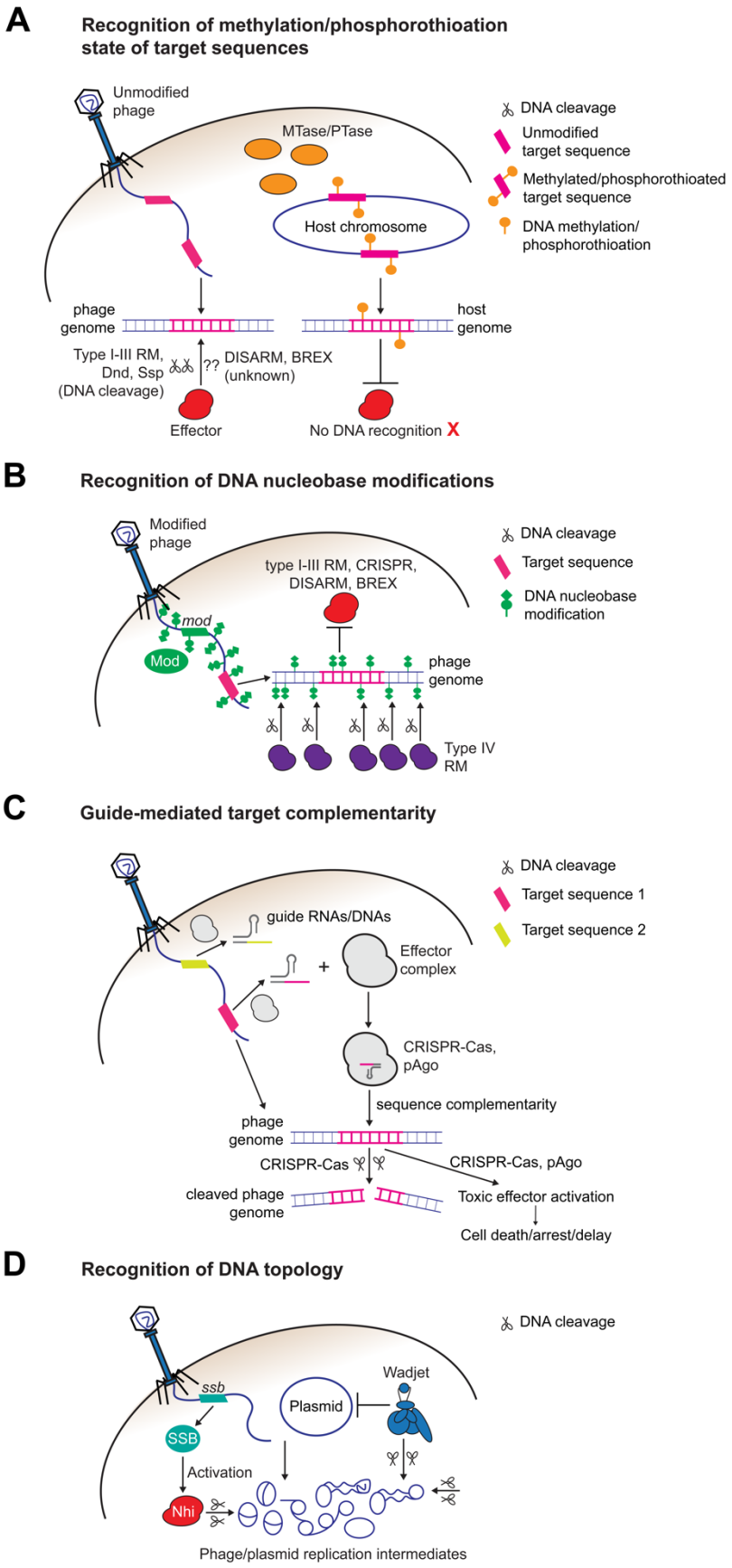

A
Recognition of methylation/phosphorothioation state of target sequences
Unmodified phage
MTase/PTase
Host chromosome
DNA cleavage
Unmodified target sequence
Methylated/phosphorothioated target sequence
DNA methylation/ phosphorothioation
phage genome
host genome
Type I-III RM, Dnd, Ssp (DNA cleavage)
??
DISARM, BREX (unknown)
Effector
No DNA recognition X
B
Recognition of DNA nucleobase modifications
Modified phage
type I-III RM, CRISPR, DISARM, BREX
mod
Mod
phage genome
Type IV RM
DNA cleavage
Target sequence
DNA nucleobase modification
C
Guide-mediated target complementarity
guide RNAs/DNAs
Effector complex
CRISPR-Cas, pAgo
sequence complementarity
phage genome
CRISPR-Cas
CRISPR-Cas, pAgo
cleaved phage genome
Toxic effector activation
Cell death/arrest/delay
DNA cleavage
Target sequence 1
Target sequence 2
D
Recognition of DNA topology
ssb
Plasmid
Wadjet
SSB
Activation
Nhi
Phage/plasmid replication intermediates
DNA cleavage

Figure 1.7 Recognition of foreign DNA by DNA-targeting bacterial defense systems

(**A**) Many DNA-targeting antiviral systems, including RM, BREX and DISARM, recognize the methylation state of their target sequences to distinguish between self and non-self DNA. The system encodes a methyltransferase (MTase) that methylates target sequences in the host chromosome, allowing recognition and targeting of foreign DNA that lacks methylation of its target sequences. Dnd and Ssp systems encode a sulfur-modification complex (PTase) that adds phosphorothioate (PT) modifications by replacing the bridging oxygen with sulfur in sugar-phosphate backbones. (**B**) Many phages carry modified nucleobases in their genomes to prevent cleavage by RM and CRISPR-Cas nucleases. Type IV RM systems recognize specific modified nucleobases and cleave modified DNA. (**C**) CRISPR-Cas and pAgo systems use short nucleic acid guides to recognize complementary target sequences within phage or plasmid DNA. These guide RNAs or DNAs are derived from the genomes of infecting phages or plasmids. While many CRISPR-Cas systems directly cleave target DNA, pAgos and some CRISPR systems mediate abortive infection upon target recognition. (**D**) A few recently discovered immune systems, specifically Nhi and Wadjet, recognize the specific topology of phage or plasmid DNA and their replication intermediates to selectively cleave foreign DNA over host DNA.

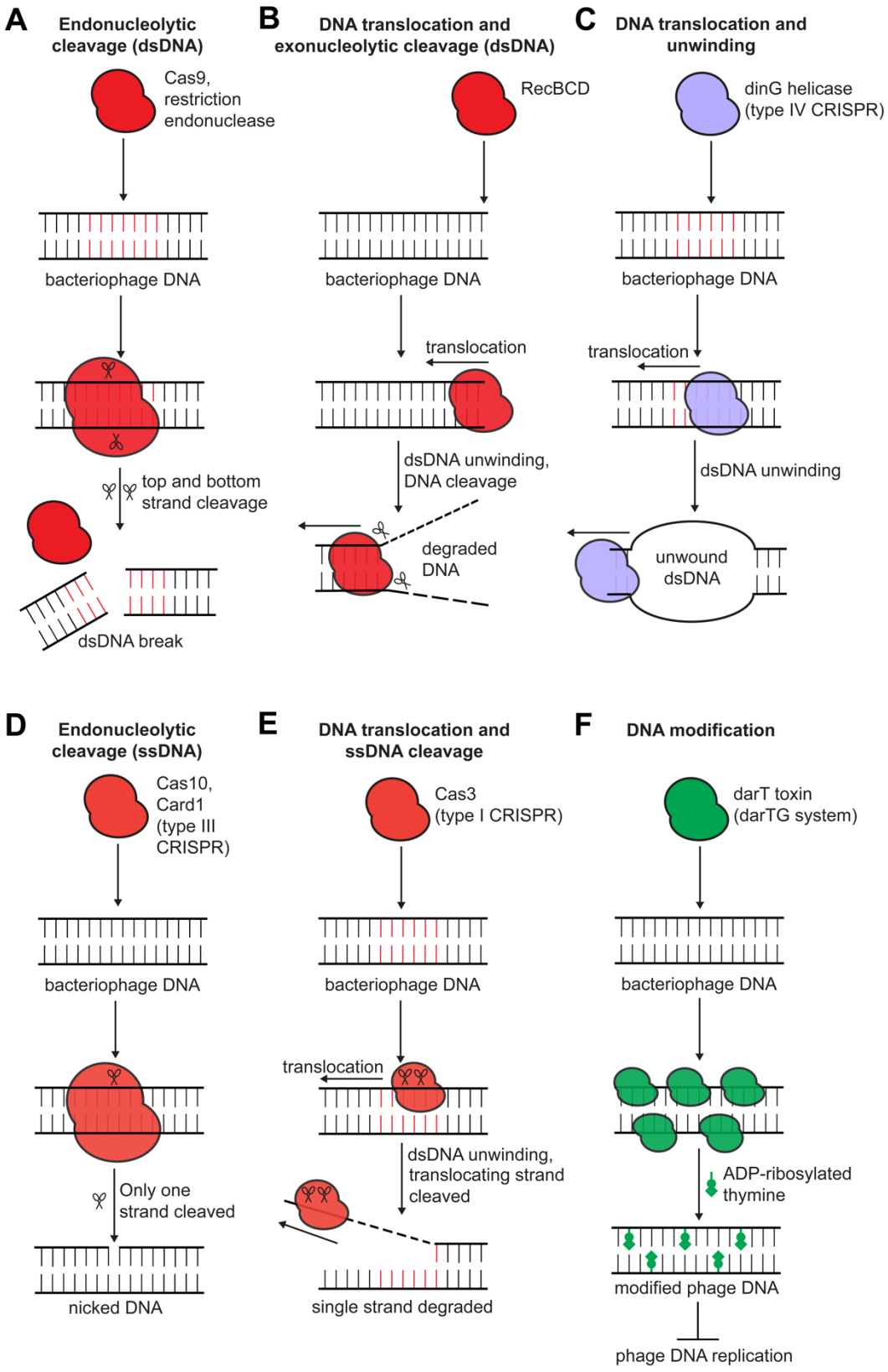

A
Endonucleolytic cleavage (dsDNA)
Cas9, restriction endonuclease
bacteriophage DNA
top and bottom strand cleavage
dsDNA break
B
DNA translocation and exonucleolytic cleavage (dsDNA)
RecBCD
bacteriophage DNA
translocation
dsDNA unwinding, DNA cleavage
degraded DNA
C
DNA translocation and unwinding
dinG helicase (type IV CRISPR)
bacteriophage DNA
translocation
dsDNA unwinding
unwound dsDNA
D
Endonucleolytic cleavage (ssDNA)
Cas10, Card1 (type III CRISPR)
bacteriophage DNA
Only one strand cleaved
nicked DNA
E
DNA translocation and ssDNA cleavage
Cas3 (type I CRISPR)
bacteriophage DNA
translocation
dsDNA unwinding, translocating strand cleaved
single strand degraded
F
DNA modification
darT toxin (darTG system)
bacteriophage DNA
ADP-ribosylated thymine
modified phage DNA
phage DNA replication

Figure 1.8 DNA damage induced by DNA-targeting bacterial defense systems

(**A**) The generation of blunt or staggered DSBs by endonucleases is the most widespread form of DNA damage generated by DNA-targeting anti-phage systems, including Cas9 and classical restriction endonucleases. (**B**) Helicase-nucleases, such as RecBCD, unwind free DNA ends while translocating along dsDNA. Exonuclease activity cleaves top and bottom strands during translocation. (**C**) The type IV CRISPR-Cas effector DinG is a helicase that unwinds and translocates along dsDNA. (**D**) The type III CRISPR-Cas effector Cas10 cleaves ssDNA and generates single strand nicks in dsDNA, while the type III CRISPR accessory effector Card1 has been shown to cleave ssDNA. (**E**) Helicase-nucleases, such as the type I CRISPR nuclease Cas3, unwind dsDNA and translocate along a single strand. ssDNA cleavage during translocation results in processive strand degradation. (**F**) The darT toxin mediates modification of DNA nucleobases, specifically the ADP-ribosylation of thymines. ADP-ribosylated genomes cannot be replicated by the viral DNA polymerase.

1.8 Restriction-modification: innate immunity targeting foreign DNA

Restriction-modification (RM) systems are the most abundant antiviral systems encoded in prokaryotes, present in 83% of bacterial genomes (Tesson et al., 2022). Minimally, they constitute two components: a DNA restriction endonuclease (R) subunit and a DNA methyltransferase (M) subunit (Loenen et al., 2014b; Tock and Dryden, 2005). Both act on the same target DNA motif, usually a 4-8 bp sequence. The methyltransferase recognizes the target sequence and adds a methyl group to a specific DNA base within the sequence. The restriction endonuclease recognizes the same target sequence and generates a DNA double-strand break (DSB) upon recognition of an unmethylated target sequence. Methylation of target sequences within the host DNA by the RM methylase prevents self-cleavage by the restriction

endonuclease and hence autoimmunity (**Fig. 1.7-A**). RM systems constitute an example of innate immunity because they only provide defense against phages and plasmids that already carry their exact target sequence and are unable to acquire resistance against invaders that lack these sequences. Consequently, phages that carry a single target sequence can evade RM immunity through mutation of this sequence. In laboratory experiments, a more common mechanism of RM evasion occurs through spurious methylation of an infecting phage genome by the RM methylase at target sequences (Maguin et al., 2022). Unmethylated genomes are rapidly eliminated by the restriction endonuclease, while methylated genomes can replicate unperturbed. This results in phage progeny with methylated genomes that are now fully resistant to RM cleavage (Bickle and Kruger, 1993; Maguin et al., 2022; Tock and Dryden, 2005). RM systems must therefore maintain a careful balance between their methyltransferase and endonuclease activities to prevent autoimmunity but still enable robust anti-phage immunity that minimizes epigenetic escape.

There are four main types of RM systems, classified according to their mode of cleavage, their protein components, and their target sequences (Tock and Dryden, 2005). Types I and III RM systems possess a helicase activity that drives translocation along DNA following target sequence recognition (Loenen et al., 2014a; Tock and Dryden, 2005). DNA translocation continues until the enzyme complex hits a physical barrier, in some cases the same enzyme translocating in the reverse direction from a separate recognition motif. Stalling of the complex triggers DNA

cleavage. As such, these systems cleave DNA non-specifically at locations away from their recognition motifs. This contrasts with type II systems, which cleave DNA within their recognition sequence, usually a 4-8 bp palindrome, or at a short specified distance away from the sequence, generating a DSB with blunt ends or short single-stranded overhangs (Tock and Dryden, 2005). Their precise cleavage activity lends to type II enzymes being widely used in molecular biology applications including cloning.

Recently, a new family of type II restriction enzymes, the R.PabI superfamily, has been described (Ishikawa et al., 2005; Miyazono et al., 2014). The R.PabI family enzymes do not possess classical endonuclease activity. Rather these enzymes are adenine DNA glycosylases with a “half-pipe” structural fold (Miyazono et al., 2014). They recognize and excise unmodified adenine nucleobases from target 5’-GTAC-3’ sites in dsDNA (Ishikawa et al., 2005; Miyazono et al., 2014). Abasic sites are generated on both strands of the palindromic target motif, which are then resolved to DSBs through enzyme- and/or host-encoded apurinic/apyrimidinic (AP) endonuclease activities that cleave the DNA sugar-phosphate backbone adjacent to an abasic site (Fukuyo et al., 2015; Zhang et al., 2017). Like the canonical type II R-M systems, the R.PabI family systems also possess a methyltransferase subunit that methylates adenines within 5’-GTAC-3’ sites to prevent cleavage of host DNA.

Type IV RM systems differ from other RM types in their recognition of target DNA (Loenen and Raleigh, 2014; Tock and Dryden, 2005). While the other systems encode both a methylase that methylates target motifs and a restriction endonuclease that cleaves unmodified DNA, type IV systems possess only a single restriction endonuclease effector that cleaves modified, typically methylated DNA (**Fig. 1.7-B**). These enzymes are referred to as modification-dependent restriction enzymes (Loenen and Raleigh, 2014) and are typified by the McrBC and Mrr systems of *E. coli* (Stewart et al., 2000; Waite-Rees et al., 1991). Type IV systems exhibit low sequence specificity and target a range of modified nucleobases, including N6-methyladenine restricted by Mrr (Heitman and Model, 1987; Waite-Rees et al., 1991), 5-methylcytosine restricted by McrBC and Mrr (Dila et al., 1990; Waite-Rees et al., 1991), 5-hydroxymethylcytosine (hmC) restricted by McrBC (Dila et al., 1990), and glucosylated hmC restricted by GmrS-GmrD (Bair and Black, 2007). Interestingly, glucosylation of hmC nucleobases, as seen in T-even phages such as T4, represents a viral “anti-restriction” strategy as this prevents cleavage of the phage DNA by McrBC in *E. coli*. The two-component type IV system GmrS-GmrD evolved in *E. coli* to enable cleavage of glucosylated T-even bacteriophage DNA (Bair and Black, 2007). In response, phage T4 expresses the GmrS-GmrD inhibitor IPI* that enables phage propagation in the presence of this system (Bair and Black, 2007; Bair et al., 2007). Taking the arms race a step further, *E. coli* has evolved a single fused GmrSD variant that is resistant to IPI* inhibition (Rifat et al., 2008). Glucosylation, GmrSD

targeting and IPI* inhibition depict a dynamic portrait of the coevolutionary arms race between a bacterium and its phages.

1.9 CRISPR-Cas: adaptive immunity that "remembers" foreign DNA

Clustered regularly interspaced short palindromic repeat (CRISPR) loci and CRISPR-associated (*cas*) genes are a family of adaptive immune systems found in 85% of archaea and 40% of bacteria (Makarova et al., 2020). These systems have been extensively studied since they were first shown to function in bacterial immunity, owing to their subsequent adoption as programmable gene editing tools. The programmable nature of CRISPR derives from the ability of these systems to adapt to foreign genetic elements and provide robust immunity against new invaders in a sequence-specific manner (Barrangou et al., 2007; Garneau et al., 2010; Jinek et al., 2012; Marraffini and Sontheimer, 2008). CRISPR loci consist of partially palindromic 30-40 bp repeat sequences separated by similarly sized "spacer" sequences. These spacers are derived from the DNA of phages and plasmids and thus represent an immunological memory of previously encountered foreign invaders (Bolotin et al., 2005; Mojica et al., 2005). CRISPR-Cas is the only known adaptive immune system found in bacteria.

CRISPR immunity can be divided into two main stages: acquisition and interference (**Fig. 1.9**) (Marraffini, 2015). In the acquisition stage, the Cas acquisition complex encoded by a subset of the *cas* genes, minimally encoding Cas1 and Cas2,

capture short fragments of DNA (protospacers) from the invading genome and incorporate these fragments as spacers in between two repeats in the CRISPR array via a process called spacer integration (McGinn and Marraffini, 2019). In the interference stage, the repeat-spacer array is transcribed and processed into short CRISPR RNAs (crRNAs). These crRNAs are used as guides by the Cas effector complex to target corresponding protospacer sequences within the genome of the invading phage or plasmid. Complementary base pairing between the guide RNA and the invader's genome triggers cleavage of the foreign nucleic acids by the Cas effector, providing robust adaptive immunity against phages and plasmids (Marraffini, 2015; Nussenzweig and Marraffini, 2020). Interestingly, the cleavage of phage genomes by RM systems has been shown to enhance CRISPR-Cas immunity through elevated spacer acquisition from DSBs generated in phage DNA (Maguin et al., 2022). This synergy between RM and CRISPR is functionally analogous to synergies between the innate and adaptive branches of the human immune system.

CRISPR-Cas systems are incredibly diverse but are broadly classified into two classes and six main types, although these are further classified into many different subtypes (Makarova et al., 2020). The two classes of systems are categorized based on the organization of their effector modules, where class I systems possess multi-subunit effector complexes and class II systems carry a single enzyme effector. Class I comprises types I, III and IV systems, where I and III are the most common CRISPR-Cas systems encoded in bacterial and archaeal genomes (Makarova et al.,

2020). The types I and IV systems target DNA (Cui et al., 2023; Redding et al., 2015), while type III systems target nascently transcribed RNA (Elmore et al., 2016; Goldberg et al., 2014; Kazlauskiene et al., 2016; Samai et al., 2015). Type III systems also encode several accessory Cas effectors, including nonspecific RNases (Jiang et al., 2016; Rostol and Marraffini, 2019b; Rostol et al., 2021), that are activated by cyclic oligonucleotide signaling generated by the primary Cas effector complex upon target RNA binding (Kazlauskiene et al., 2017; Niewoehner et al., 2017). These accessory effectors induce cellular dormancy or abortive infection in response to activation (Rostol and Marraffini, 2019b; Rostol et al., 2021).

Class II systems encode large single subunit effector enzymes, for example the Cas9 enzyme from *Streptococcus pyogenes* that is widely used for programmable, sequence-specific gene editing (Cong et al., 2013; Jinek et al., 2012). Among the systems in this class, type II and V enzymes, Cas9 and Cas12, respectively, recognize DNA and mediate targeted DNA cleavage (Nussenzweig and Marraffini, 2020; Sternberg et al., 2014; Zetsche et al., 2015). Type VI systems target RNA and trigger nonspecific RNA cleavage upon target RNA binding to induce cellular dormancy (Abudayyeh et al., 2016; Meeske et al., 2019). In this thesis, I will focus on the well-studied type II-A and type I-E CRISPR Cas-systems of *S. pyogenes* and *E. coli*, respectively, as experimental models for DNA-targeting CRISPR-Cas immunity in bacteria.

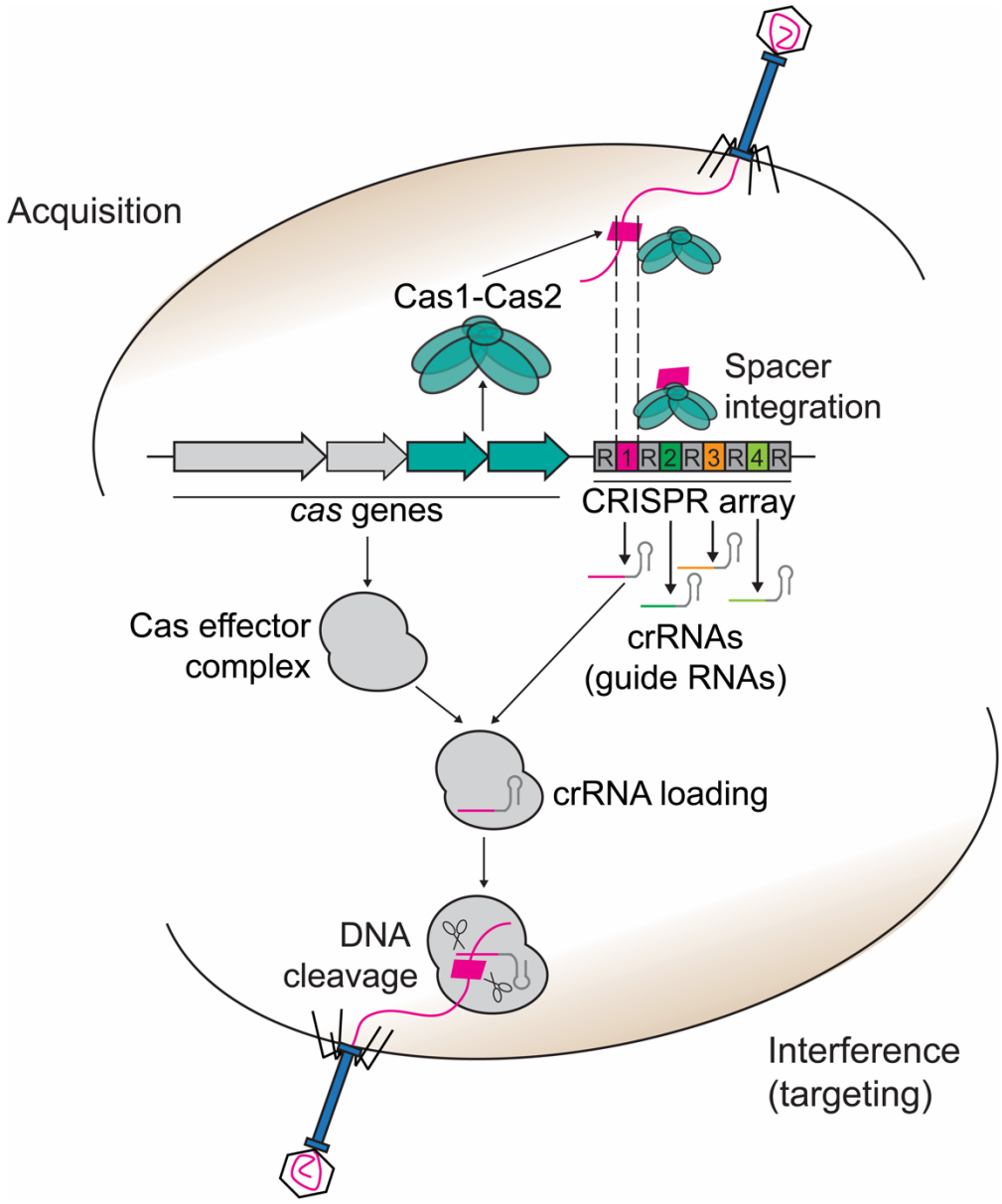


Figure 1.9 CRISPR-Cas immunity in bacteria

CRISPR-Cas immunity is divided into two stages: acquisition and interference (targeting). During acquisition, a short fragment of the invading phage genome known as a protospacer is captured by the Cas1-Cas2 complex and integrated into the CRISPR array as a new spacer in between two repeats in the array. During interference, also referred to as targeting, the CRISPR array is transcribed and processed to generate guide RNAs used by the Cas effector complex to recognize and bind complementary target sequences (protospacers) in foreign DNA. Upon target recognition, the target DNA is typically cleaved by the Cas effector complex, although in certain cases, the effector complex may activate accessory effectors that mediate abortive infection through cell arrest or death. Figure adapted from Marraffini (2015) and reproduced with permission from Springer Nature.

1.10 The type II-A CRISPR-Cas system of *Streptococcus pyogenes*

CRISPR-Cas immunity has been studied extensively in the Marraffini lab using the Gram-positive *S. pyogenes* SF370 type II-A CRISPR-Cas system (Heler et al., 2015; Maguin et al., 2022; McGinn and Marraffini, 2016; Modell et al., 2017; Nussenzweig et al., 2019; Varble et al., 2021). This system encodes the Cas9 enzyme used for genome editing in eukaryotic cells owing to its ability to generate a programmable, site-specific DSB (Cong et al., 2013; Jinek et al., 2012). The *S. pyogenes* system is a good laboratory model because of its simplicity: the locus only encodes *tracr* and four *cas* genes, upstream of an AT-rich leader sequence that precedes the CRISPR repeat-spacer array (Heler et al., 2015; McGinn and Marraffini, 2019) (**Fig. 1.10**). While all four Cas proteins (Cas1, Cas2, Csn2 and Cas9) participate in spacer acquisition (Heler et al., 2015), interference with a preexisting spacer derived from the CRISPR array requires only Cas9 and a trans-acting accessory RNA called tracrRNA (Chylinski et al., 2013; Deltcheva et al., 2011; Jinek et al., 2012). Cas9 loaded with tracrRNA base pairs with a long precursor crRNA transcribed from the CRISPR array through complementarity between tracrRNA and the CRISPR repeat sequence. RNase III from the host cleaves the precursor crRNA at the repeats to give short, mature crRNAs that function as guide RNAs for the Cas9-tracrRNA complex (Deltcheva et al., 2011). Together, this Cas9-crRNA-tracrRNA complex is the type II-A CRISPR interference effector (**Fig. 1.10**). The Cas9 effector complex uses the crRNA to identify the target sequence (protospacer), located immediately upstream of a protospacer adjacent motif (PAM) (Sternberg et

al., 2014). For the *S. pyogenes* SF370 type II-A system, the PAM is 5'-NGG-3', directly downstream of the 30 bp protospacer sequence (Jinek et al., 2012). A PAM binding domain within Cas9 binds the PAM (Anders et al., 2014; Jinek et al., 2014; Nishimasu et al., 2014), melting dsDNA immediately upstream in the 8-12 bp "seed" region of the protospacer (Anders et al., 2014; Jiang et al., 2015; Sternberg et al., 2014). Binding of the crRNA to the seed sequence on the target strand activates DNA cleavage by the RuvC and HNH endonuclease domains of Cas9, which cleave one strand each of the protospacer to generate a DSB (Gasiunas et al., 2012; Jinek et al., 2012; Sternberg et al., 2015). Mutations in the RuvC and HNH domains, D10A and H840A, respectively, render Cas9 "nuclease dead". The D10A, H840A mutant Cas9, referred to as dCas9, binds the protospacer sequence without cleaving it, forming a roadblock that stalls DNA/RNA polymerases (Bikard et al., 2013; Qi et al., 2013).

Following Cas9 cleavage, host nucleases are expected to degrade the cleaved DNA, although this hypothesis remains to be demonstrated. Degradation by host nucleases is thought to facilitate immunity by promoting clearance of phage or plasmid DNA. To evade Cas9 targeting, phages often encode anti-CRISPR proteins that bind Cas9 and impede its tracrRNA binding, target recognition and/or DNA cleavage activities (Davidson et al., 2020). Additionally, phages lacking anti-CRISPRs can mutate their Cas9 target sequence. CRISPR escape mutations arise as single point mutations in the PAM which prevent target recognition, as point

mutations in the seed sequence that impair cRNA binding to the target, or as partial or full deletions of the Cas9 protospacer and/or PAM sequence (Deveau et al., 2008; Nussenzweig et al., 2019; Pyenson et al., 2017).

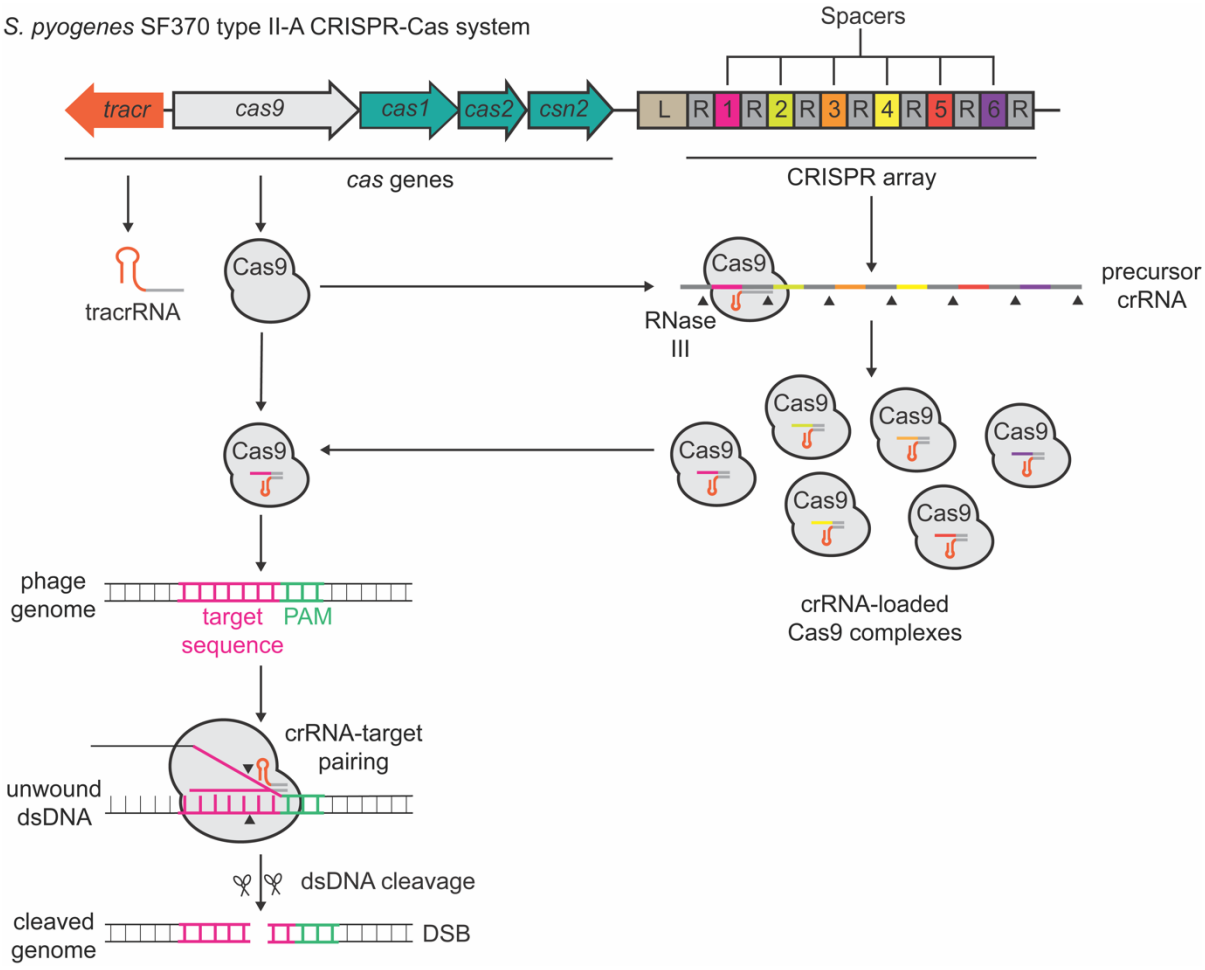


Figure 1.10 Cas9 targeting in the *S. pyogenes* type II-A CRISPR-Cas system

In the *S. pyogenes* type II-A system, Cas9 and tracrRNA are required for targeting. The CRISPR array is transcribed to give a long precursor crRNA. RNase III cleaves the precursor RNA to give mature crRNAs used by the Cas9:tracrRNA complex. The crRNA-loaded Cas9 effector complex performs target search through PAM recognition. Upon PAM recognition, Cas9 unwinds the upstream target sequence, allowing complementary base pairing between the target sequence and the crRNA. Upon target complementarity, Cas9 cleaves the top and bottom strands of the target sequence to generate a DSB.

1.11 The type I-E CRISPR-Cas system of *Escherichia coli*

Unlike type II systems, type I CRISPR-Cas systems use a large multi-subunit interference complex called Cascade that recognizes target DNA and recruits the trans-acting nuclease-helicase Cas3 upon target binding (Brouns et al., 2008; Nussenzweig and Marraffini, 2020; Westra et al., 2012; Zhao et al., 2014) (**Fig. 1.11**). These systems are the most widespread in nature (Makarova et al., 2020) and are typified by the type I-E system encoded by *E. coli* K-12. In the *E. coli* Cascade complex, the Cas6 subunit cleaves the long precursor crRNA, transcribed from the type I-E CRISPR array, to give short, processed crRNAs that are used as guides by Cascade (Brouns et al., 2008). The Cas8 subunit (previously CasA or Cse1) of Cascade is the PAM binding domain of the effector (Sashital et al., 2012), recognizing a promiscuous 5'-AWG-3' PAM immediately upstream of the 32 bp protospacer sequence (Westra et al., 2012). PAM recognition is less stringent than the strict PAM selectivity observed in the *S. pyogenes* type II-A system (Fu et al., 2017; Heler et al., 2015; Jiang et al., 2013). The first 8 bp of the protospacer, immediately downstream of the PAM, comprises the seed sequence (Semenova et al., 2011; Wiedenheft et al., 2011). As with the type II system, mutations in the PAM or seed sequence prevent target recognition and crRNA binding, thereby enabling viral evasion of immunity (Semenova et al., 2011). Upon crRNA binding to target DNA, the Cas8 subunit of Cascade recruits the trans-acting Cas effector Cas3, which functions as the nuclease in this system (Hochstrasser et al., 2014; Huo et al., 2014; Sinkunas et al., 2011; Sinkunas et al., 2013; Westra et al., 2012). Cas3 is also an

ATP-dependent helicase (Huo et al., 2014; Sinkunas et al., 2011; Westra et al., 2012), which first cuts both target and non-target strands and then translocates along the non-target strand and mediates DNA degradation of this strand by generating breaks in the unwound ssDNA during translocation (Mulepati and Bailey, 2013; Redding et al., 2015; Sinkunas et al., 2013). Cas3-mediated degradation is thought to ultimately result in full dsDNA degradation in vivo, but the mechanisms driving this are currently unclear (Nussenzweig and Marraffini, 2020; Sinkunas et al., 2013). The type I-E system in *E. coli* is transcriptionally repressed by the global DNA-binding transcriptional regulator H-NS (Pul et al., 2010; Westra et al., 2010). Since this system is inactive under laboratory conditions, in vivo experimental studies using this system have required placing the *cas* genes and the CRISPR array under the artificial control of inducible promoters.

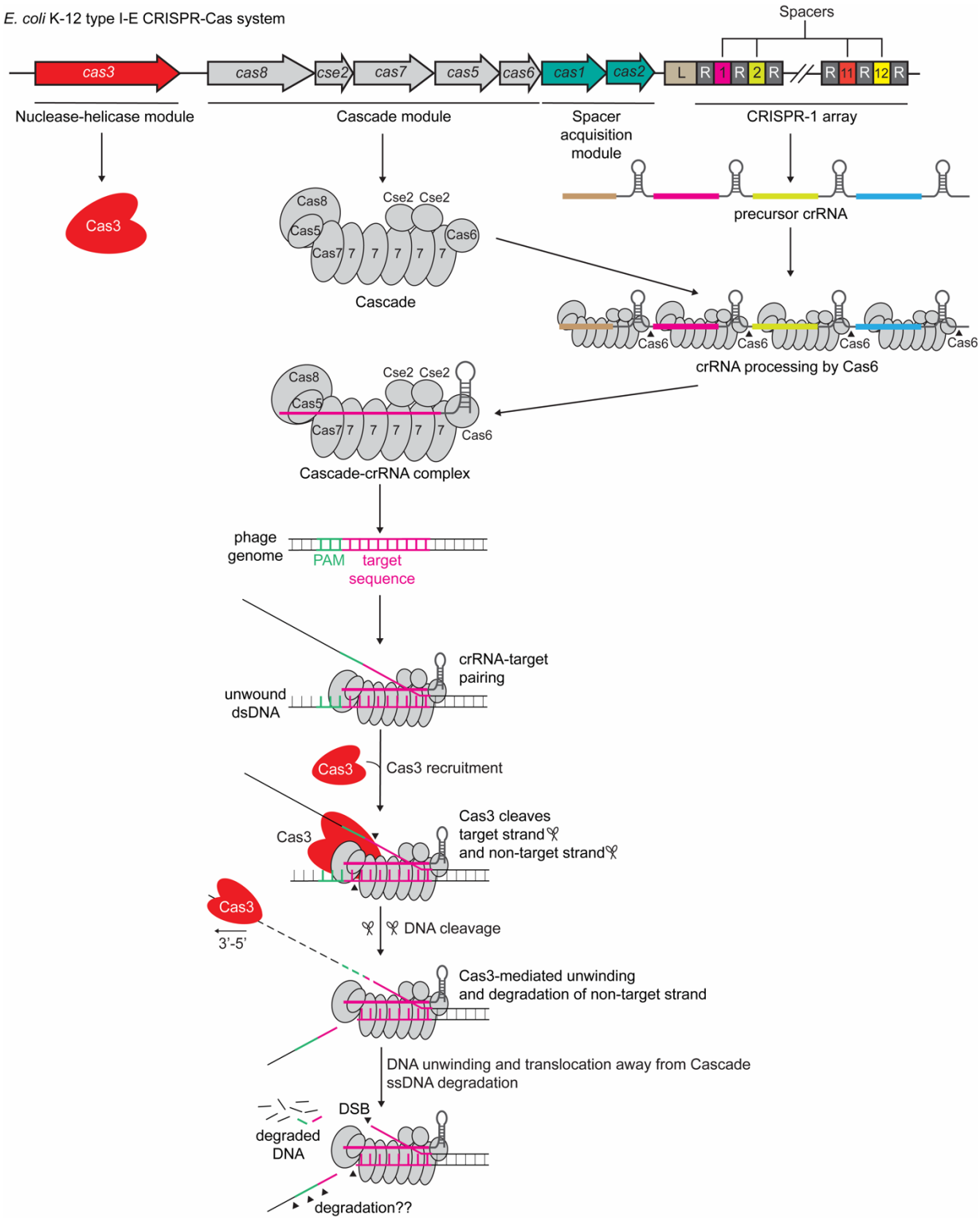

E. coli K-12 type I-E CRISPR-Cas system
Spacers
cas3
cas8
cse2
cas7
cas5
cas6
cas1
cas2
L R 1 R 2 R R 11 R 12 R
Nuclease-helicase module
Cascade module
Spacer acquisition module
CRISPR-1 array
Cas3
Cas8
Cse2 Cse2
Cas5
Cas7 7 7 7 7 7
Cas6
Cascade
precursor crRNA
Cas6
crRNA processing by Cas6
Cascade-crRNA complex
phage genome
PAM
target sequence
crRNA-target pairing
unwound dsDNA
Cas3 recruitment
Cas3 cleaves target strand and non-target strand
DNA cleavage
3'-5'
Cas3-mediated unwinding and degradation of non-target strand
DNA unwinding and translocation away from Cascade
ssDNA degradation
DSB
degraded DNA
degradation??

Figure 1.11 Cascade-Cas3 targeting in the *E. coli* type I-E CRISPR-Cas system

In the *E. coli* type I-E system, the large multi-subunit complex Cascade and the nuclease-helicase Cas3 execute DNA targeting. The CRISPR array is transcribed to generate a long precursor crRNA. Cas6 in the Cascade complex cleaves the precursor RNA to give mature crRNAs used by Cascade. The crRNA-loaded Cascade effector complex performs target search through PAM recognition. Upon PAM recognition by Cas8, Cascade unwinds the target sequence, allowing complementary base pairing between the target sequence and the crRNA. Upon target complementarity, Cascade, bound to the target sequence, recruits the Cas3 nuclease-helicase through its Cas8 subunit. Cas3 cuts both target and non-target strands and then unidirectionally translocates along the non-target strand in the 3'-5' direction, away from Cascade, generating ssDNA breaks in this strand during translocation and driving strand degradation. It is unclear whether the cleaved overhang strand (target strand) is degraded further in vivo.

1.12 Phage strategies to counter bacterial immunity

As outlined already, bacteria employ a wide variety of immune mechanisms to thwart phages (Georjon and Bernheim, 2023; Hampton et al., 2020; Samson et al., 2013). To enable successful infections, phages have evolved a myriad of counter strategies that evade, inhibit, or minimize these bacterial defenses (**Fig. 1.12**). The simplest of these mechanisms involves mutation of the phage genome so that the target of bacterial immunity is genetically altered or deleted and can therefore no longer be recognized by the host (**Fig. 1.12-A**). In the case of CRISPR-Cas immunity, escape mutations map to the PAM or protospacer sequence targeted by the guide RNA (Deveau et al., 2008; Nussenzweig et al., 2019; Pyenson et al., 2017; Semenova et al., 2011). For abortive infection systems that sense phage effector-triggered activities or viral PAMPs, mutations in the genes encoding virulence

effectors or the phage structural proteins that serve as PAMPs provide a route to evading immune surveillance (Stokar-Avihail et al., 2023). In retron systems that monitor the activity of the bacterial recombinase-nuclease RecBCD, immune activation is triggered through RecBCD inhibition by a phage effector protein, for example the Gam protein expressed by phage λ (Millman et al., 2020a; Stokar-Avihail et al., 2023). Here, mutation or deletion of Gam leads to phage evasion of retron-mediated immunity. Another common route for phages to evade host defenses involves epigenetic escape, through the gain or loss of DNA modifications. This is particularly relevant for RM and RM-like systems such as BREX and DISARM that rely on differential epigenetic states, specifically methylation, to distinguish foreign DNA from self DNA (Bickle and Kruger, 1993; Gordeeva et al., 2018; Ofir et al., 2018; Samson et al., 2013; Tock and Dryden, 2005). During infection, the spurious methylation of phage genomes prevents cleavage by host restriction endonucleases, enabling phage replication and the rise of epigenetically modified escaper viruses now resistant to immunity (Bickle and Kruger, 1993; Maguin et al., 2022) (**Fig. 1.12-B**). Indeed, many phages encode their own methyltransferases and other DNA modifying enzymes that introduce varied nucleobase modifications that serve an "anti-restriction" function – they prevent recognition or cleavage by restriction and CRISPR-Cas endonucleases (Bryson et al., 2015; Kruger and Bickle, 1983; Liu et al., 2020b; Maffei et al., 2021; Samson et al., 2013; Vlot et al., 2018; Warren, 1980) (**Fig. 1.12-C**). This is best exemplified by the "hypermodification" strategies enacted by T-even phages such as T4 (Kruger and Bickle, 1983; Lehman and Pratt, 1960).

In many cases, phages genetically encode proteins that directly inhibit bacterial immune systems (**Fig. 1.12-D**). These genes are typically clustered together in hotspots in the phage genome, termed anti-defense clusters (Pinilla-Redondo et al., 2020). This is the case with anti-CRISPR (Acr) proteins, which are encoded in *acr* loci in phage genomes (Bondy-Denomy et al., 2013; Pawluk et al., 2018; Pawluk et al., 2016b; Pinilla-Redondo et al., 2020). Acrs typically bind CRISPR-Cas effector complexes, preventing their activity by disrupting crRNA binding, target recognition, complex formation and/or nuclease activity (Bondy-Denomy et al., 2015; Davidson et al., 2020). Other examples of anti-defense clusters include the *ip1* gene locus in T-even phage genomes, which carry variable IP proteins (Rifat et al., 2008). The IPI* protein, expressed by phage T4, inhibits the type IV RM system GmrS-GmrD (Bair and Black, 2007; Bair et al., 2007). T-even phages also encode the anti-DarT factor AdfA, which inhibits the ADP-ribosylating DarT toxin of DarTG toxin-antitoxin systems that mediate anti-phage defense (LeRoux et al., 2022). Some phage-encoded inhibitors are DNA mimics that inhibit host nuclease-based systems. For example, the phage T7 anti-restriction protein Ocr is a DNA mimic that inhibits RM and BREX systems (Atanasiu et al., 2002; Isaev et al., 2020; Walkinshaw et al., 2002), as are the phage T4 anti-restriction protein Arn (Ho et al., 2014) and the phage λ Gam protein that inhibits RecBCD (Wilkinson et al., 2016). The PARIS immune system triggers abortive infection upon detection of Ocr (Rousset et al., 2022), showcasing the evolution of new layers of bacterial immunity that are necessary to respond to phages' sophisticated counter-immune strategies.

While many phage-encoded anti-defense proteins are physical inhibitors of immune complexes, some are enzymes that interfere with critical signaling molecules that drive abortive immunity. Anti-defense proteins of CBASS, Pycsar, Thoeris and type III CRISPR-Cas systems disrupt the cyclic second messenger molecules produced by these systems in response to phage infection (**Fig. 1.12-E**). The anti-CBASS protein Acb1 and the anti-Pycsar protein Apyc1 cleave the cyclic nucleotides produced by CBASS and Pycsar cyclases, respectively (Hobbs et al., 2022), while anti-CRISPR viral ring nucleases cleave cyclic oligonucleotides produced by type III CRISPR-Cas systems (Athukoralage et al., 2020). In contrast, the anti-CBASS protein Acb2 and the anti-Thoeris protein Tad1 are viral sponges that bind and sequester the cyclic signaling molecules produced by CBASS and Thoeris cyclases, respectively, preventing effector activation upon phage infection (Huiting et al., 2023; Leavitt et al., 2022). Collectively, all these examples illustrate the rich plethora of viral immune evasion strategies that accompany the diverse array of known bacterial defenses.

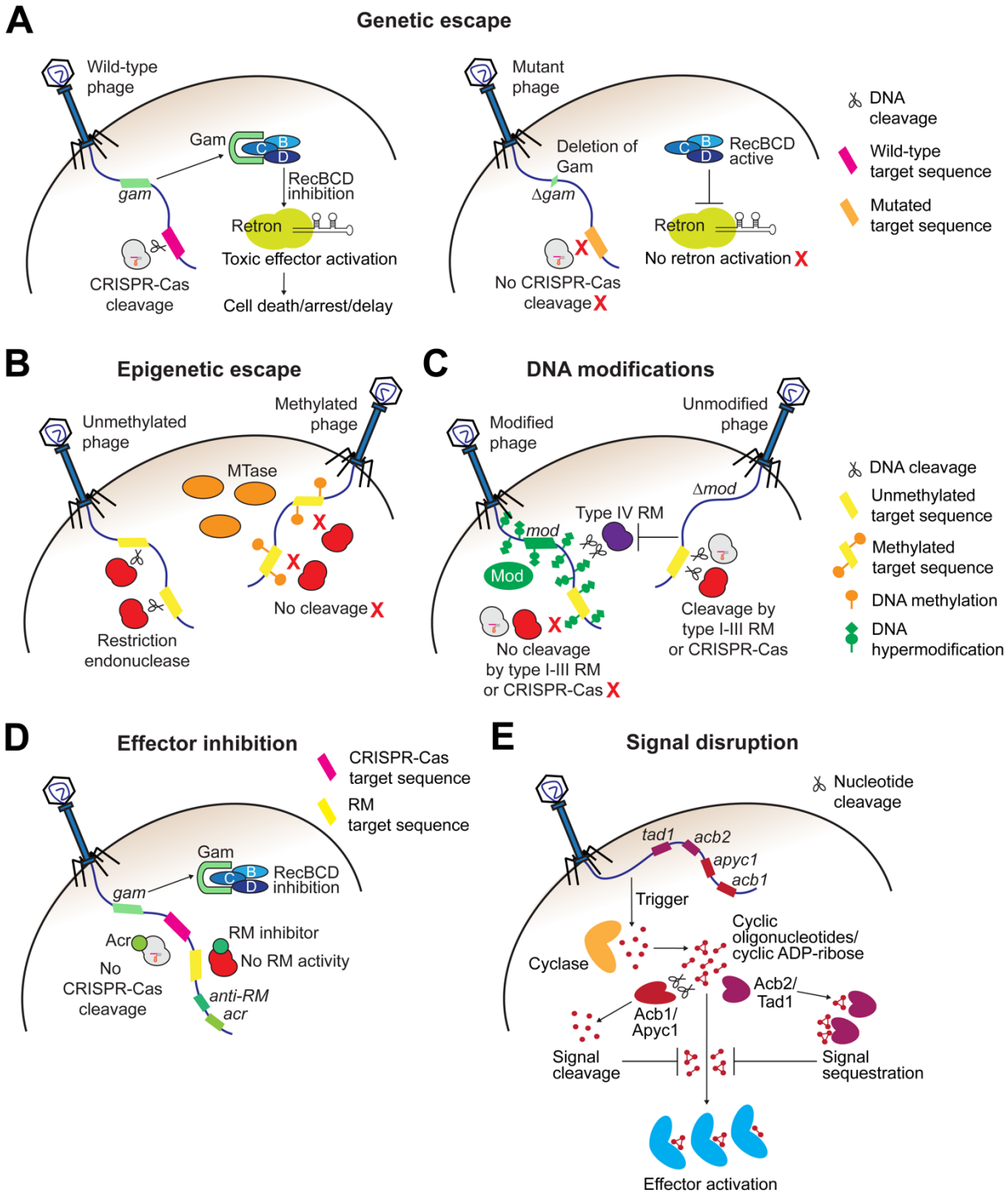

A
Genetic escape
Wild-type phage
Gam
gam
RecBCD inhibition
Retron
Toxic effector activation
CRISPR-Cas cleavage
Cell death/arrest/delay
Mutant phage
Deletion of Gam
Δgam
RecBCD active
Retron
No retron activation
No CRISPR-Cas cleavage
DNA cleavage
Wild-type target sequence
Mutated target sequence
B
Epigenetic escape
Unmethylated phage
Methylated phage
MTase
No cleavage
Restriction endonuclease
C
DNA modifications
Modified phage
Unmodified phage
Δmod
mod
Type IV RM
Mod
Cleavage by type I-III RM or CRISPR-Cas
No cleavage by type I-III RM or CRISPR-Cas
DNA cleavage
Unmethylated target sequence
Methylated target sequence
DNA methylation
DNA hypermodification
D
Effector inhibition
CRISPR-Cas target sequence
RM target sequence
Gam
gam
RecBCD inhibition
Acr
No CRISPR-Cas cleavage
RM inhibitor
No RM activity
anti-RM acr
E
Signal disruption
Nucleotide cleavage
tad1
acb2
apyc1
acb1
Trigger
Cyclase
Cyclic oligonucleotides/ cyclic ADP-ribose
Acb2/ Tad1
Acb1/ Apyc1
Signal cleavage
Signal sequestration
Effector activation

Figure 1.12 Diversity of viral counter-immune strategies

(**A**) Genetic escape is a simple and common strategy of immune evasion. Phages can mutate target sequences or delete or modify genes whose encoded proteins activate defense systems. Genetic mutation renders the phage unsusceptible or less susceptible to immune targeting. (**B**) Epigenetic escape of defense systems such as RM entails epigenetic modification of phage genomes such that the defense system can no longer distinguish foreign DNA from self DNA. With RM systems, this typically involves spurious methylation of the phage genome by the RM methyltransferase such that the methylated phage genome is no longer recognized by restriction endonucleases. (**C**) DNA nucleobase modifications, often encompassing multiple layers of modification referred to as “hypermodification”, prevent cleavage by host defense endonucleases and thus constitute an “anti-restriction” mechanism. Type IV RM systems cleave modified DNA, so phages that are targeted by type IV RM in their hosts can delete their DNA-modifying genes to lose DNA modifications and genetically escape type IV RM targeting. (**D**) Many phages encode proteins that directly inhibit host immune systems, including anti-CRISPR proteins that inhibit CRISPR-Cas effectors, Gam that inhibits the host recombinase-nuclease RecBCD, and anti-RM proteins such as Ocr encoded by phage T7 and IPI* encoded by phage T4. These inhibitors directly bind target complexes, preventing their function. (**E**) Some phages encode enzymes that disrupt immune signaling molecules produced by abortive infection systems such as CBASS, Pycsar and Thoeris. The anti-CBASS protein Acb1 and the anti-Pycsar protein Apyc1 cleave the cyclic nucleotides produced by CBASS and Pycsar cyclases respectively. The anti-Thoeris protein Tad1 and the anti-CBASS protein Acb2 act as sponges that sequester their target immune signaling molecules. In each case, the immune signaling is disrupted to block downstream effector activation. Figure modified from Georjon and Bernheim (2023) and reproduced with permission from Springer Nature.

1.13 Anti-CRISPRs: direct inhibition of CRISPR-Cas effectors

Genetic escape through target mutation does not always offer a viable route for evasion of CRISPR-Cas immunity because phage genomes are compact (Brüssow and Hendrix, 2002), so any mutation has a decent probability of disrupting the function of one or more critical phage proteins. Therefore, many phages encode

direct inhibitors of CRISPR-Cas effectors. These anti-CRISPR (Acr) proteins interrupt different steps of the CRISPR interference process by directly binding subunits of Cas effector complexes (Davidson et al., 2020; Pawluk et al., 2018). Acrs are typically expressed very early during phage infection, encoded in *acr* loci that are part of the phage's immediate early genes, which are the first to switch on upon injection (Bondy-Denomy et al., 2013; Pinilla-Redondo et al., 2020; Stanley et al., 2019; Varble et al., 2021). Bioinformatics approaches involving guilt-by-association with *acr* clusters (Marino et al., 2018; Pawluk et al., 2016a; Pawluk et al., 2016b; Pinilla-Redondo et al., 2020) and scanning genomic databases for the presence of self-targeting spacers (Rauch et al., 2017; Watters et al., 2018), as well as CRISPR self-targeting functional screens (Uribe et al., 2019), have uncovered several different Acr proteins whose targets span almost all types of CRISPR-Cas systems (Davidson et al., 2020). Acr proteins are specific to the subtype of CRISPR that they inhibit and are therefore named accordingly (Bondy-Denomy et al., 2018; Davidson et al., 2020). Inhibitors are often ortholog-specific, although broad-spectrum inhibitors have been reported (Harrington et al., 2017; Knott et al., 2019; Marino et al., 2018; Watters et al., 2018). Acrs of type I and II CRISPR-Cas systems are the most widely reported (Davidson et al., 2020) (**Fig. 1.13**). For type I systems, most reported Acrs prevent target recognition by Cascade (Bondy-Denomy et al., 2015; Chowdhury et al., 2017; Guo et al., 2017; Peng et al., 2017) while some disable Cas3-mediated DNA cleavage (Bondy-Denomy et al., 2015; Pawluk et al., 2017; Wang et al., 2016) (**Fig. 1.13-A**). For type II systems, Acrs prevent the loading of crRNAs onto Cas9

(Thavalingam et al., 2019; Zhu et al., 2019), inhibit target recognition by Cas9 (Dong et al., 2017; Harrington et al., 2017; Jiang et al., 2019; Lee et al., 2018; Liu et al., 2019a; Shin et al., 2017; Yang and Patel, 2017; Zhu et al., 2019), and inhibit Cas9 nuclease activity (Pawluk et al., 2016a; Zhu et al., 2023) (**Fig. 1.13-B**). Acr-mediated inhibition has been shown to be dose-dependent, where a sufficiently high multiplicity of phage genomes needs to be present and expressing Acr for efficient host immunosuppression (Borges et al., 2018; Landsberger et al., 2018) (**Fig. 1.14**). With only partial CRISPR-Cas inhibition at low doses, Acrs may synergize with other viral counter-immune strategies, weakening CRISPR-Cas targeting and buying time for the phage to escape through target mutation.

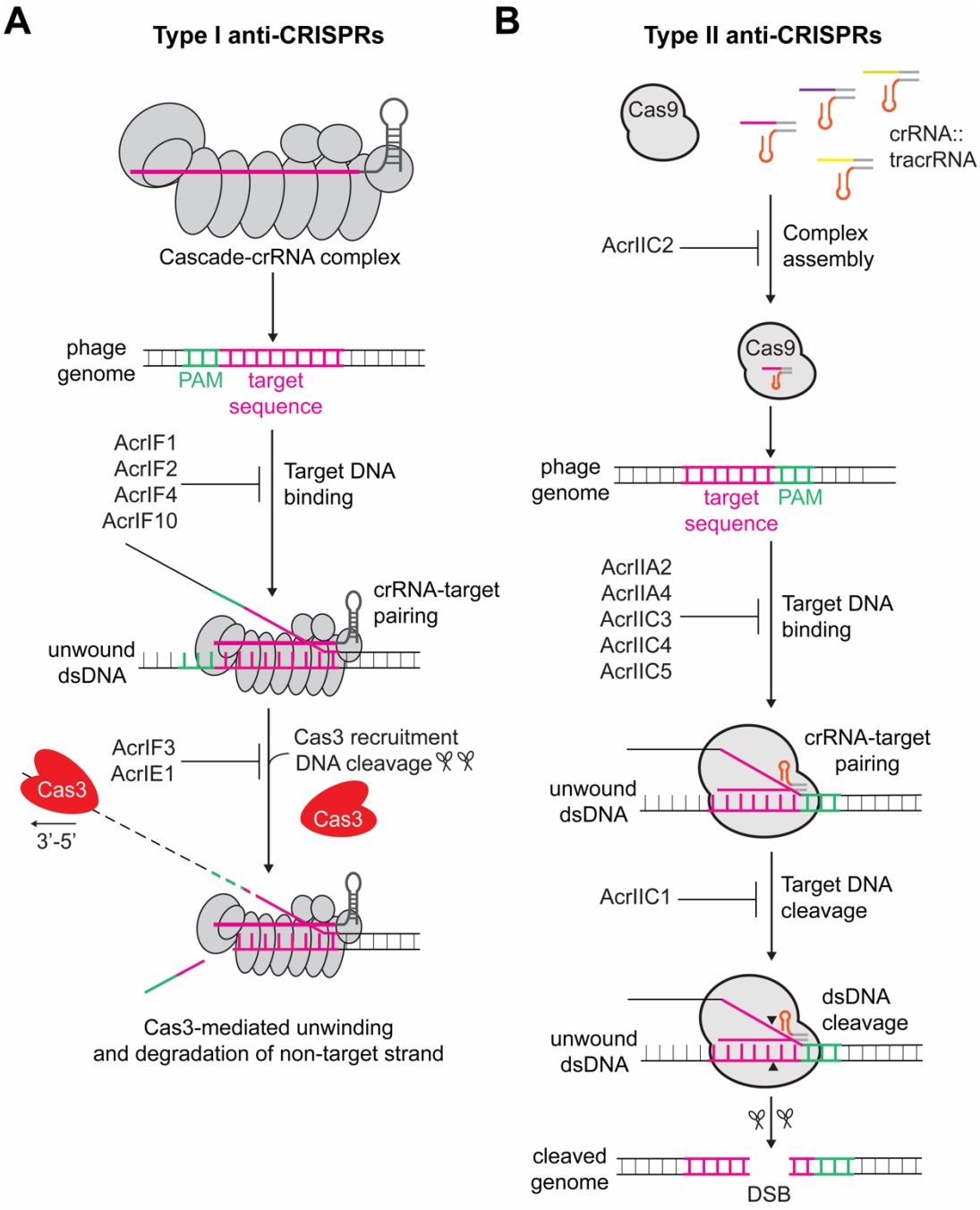


Figure 1.13 Mechanisms of anti-CRISPR inhibition of type I and II CRISPR-Cas effector complexes

(**A**) Acrs of type I CRISPR-Cas systems prevent the Cascade complex from interacting with DNA or disable Cas3 to prevent target cleavage. (**B**) Acrs of type II CRISPR-Cas systems inhibit crRNA loading onto Cas9 to disrupt complex assembly, prevent the Cas9 effector complex from recognizing target DNA, and inhibit Cas9 nuclease activity to prevent target cleavage. Figure adapted from Davidson et al. (2020) and reproduced with permission from ANNUAL REVIEWS.

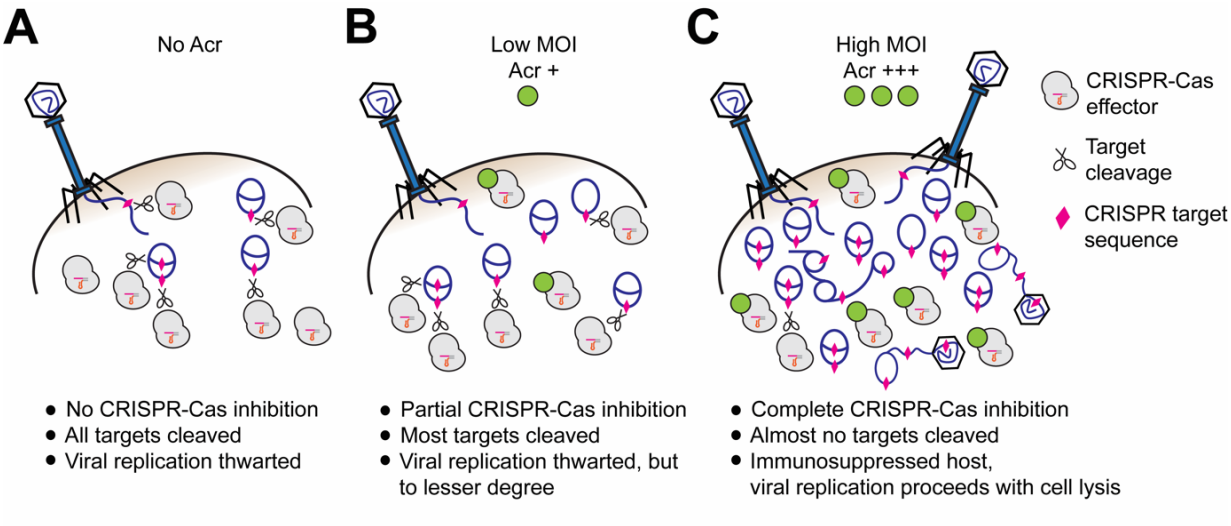


Figure 1.14 Immunosuppression of CRISPR-Cas requires a high dose of phage-encoded anti-CRISPRs

(**A**) In the absence of Acr, there is no CRISPR-Cas inhibition and Cas effector complexes cleave all target sequences. Viral replication is thwarted. (**B**) At low infection doses, low intracellular concentration of Acrs inhibits a few but not all active CRISPR-Cas complexes. Most target sequences are cleaved, and viral replication is still thwarted, albeit to a lesser degree than in the absence of Acr. (**C**) At high infection doses, there is a buildup of high intracellular Acr concentrations leading to near total inhibition of the host's CRISPR-Cas complexes. The host is immunosuppressed, allowing a phage that now infects the immunosuppressed host to replicate freely and overcome CRISPR-Cas immunity.

1.14 DNA modifications: a phage anti-restriction strategy

Many phages carry atypical nucleobases (Warren, 1980; Weigele and Raleigh, 2016). These protect viral DNA from recognition and cleavage by restriction enzymes, CRISRP-Cas, and other host nucleases (Bryson et al., 2015; Kruger and Bickle, 1983; Liu et al., 2020b; Maffei et al., 2021; Olsen et al., 2023; Samson et al., 2013; Vlot et al., 2018). It is well documented that T-even coliphages "hypermodify" their nucleobases, which serves an anti-restriction function (Kruger and Bickle, 1983;

Lehman and Pratt, 1960). All of the cytosine nucleobases in T-even phage genomes are replaced by 5-hydroxymethylcytosine (hmC) (Wyatt and Cohen, 1953) (**Fig. 1.15**). This occurs prior to DNA synthesis, at the level of the cellular nucleotide pool. In phage T4, this is driven by two genes, *gp42* and *gp56*, which encode a dCMP hydroxymethylase and a dCTPase, respectively (Carlson and Wiberg, 1983; Chiu et al., 1976; Lamm et al., 1988; Wiberg, 1967). The Gp42 dCMP hydroxymethylase generates hmC mononucleotides while the Gp56 dCTPase degrades deoxycytidine nucleotides in the cytosol. Additional T4-encoded enzymes, such as the dC-specific ssDNA endonuclease denB (Carlson and Wiberg, 1983; Hirano et al., 2006) and the dC-specific premature transcriptional terminator Alc (Drivdahl and Kutter, 1990; Severinov et al., 1994), select against dC-containing phage and host DNA, by degrading any dC-containing DNA and inhibiting its transcription. The collective action of all these phage enzymes gives a T4 phage genome where all cytosine nucleobases are replaced by 5-hydroxymethylcytosine (Bryson et al., 2015).

While hmC blocks cleavage by some restriction endonucleases, many type IV RM systems and CRISPR-Cas endonucleases recognize and cleave hmC-modified DNA. Therefore, T4 and other T-even phages further modify their hmC nucleobases by glycosylating them with diverse glucose adducts (Kuno and Lehman, 1962; Lehman and Pratt, 1960) (**Fig. 1.15**). The most common of these “hypermodified” nucleobases is alpha-glucosyl-hmC (Lehman and Pratt, 1960), generated by the T-even phage enzyme alpha-glucosyltransferase (a-GT). During replication, a-GT

associates with the phage replisome and alpha-glucosylates newly synthesized hmC nucleobases during DNA replication (Sommer et al., 2004). a-GT is unable to modify two hmC nucleobases right next to each other, resulting in a phage genome in which only 70-75% of the hmC nucleobases are alpha-glucosylated (de Waard et al., 1967; Lunt and Newton, 1965). In phage T4, the remaining ~30% of hmC nucleobases are glucosylated by a second enzyme, beta-glucosyltransferase (b-GT), which generates beta-glucosyl-hmC nucleobases (Georgopoulos and Revel, 1971; Lehman and Pratt, 1960) (**Fig. 1.15-AC**). Phages T2 and T6 lack b-GT, so ~25% of the hmC nucleobases in their genomes are unmodified (Lehman and Pratt, 1960; Lunt and Newton, 1965). Instead T2 and T6 encode the enzyme beta-alpha glucosyltransferase (ba-GT), that adds a second glucose in beta linkage to alpha-glucosyl-hmC nucleobases, to varying degrees in T2 and T6 genomes. The doubly glucosylated hmC nucleobase is called gentiobiosyl-hmC (Kuno and Lehman, 1962) (**Fig. 1.15-BC**). While 70% of the hmC nucleobases in phage T2 are alpha-glucosyl-hmC and 5% are gentiobiosyl-hmC, 72% in T6 are gentiobiosyl-hmC and only 3% are alpha-glucosyl-hmC (Lehman and Pratt, 1960). Glycosylation of hmC protects the phage DNA against cleavage by many restriction endonucleases, CRISPR-Cas endonucleases such as Cas9, and type IV RM systems such as McrBC (Bryson et al., 2015; Kruger and Bickle, 1983; Liu et al., 2020b; Maffei et al., 2021; Samson et al., 2013; Vlot et al., 2018). The type IV RM system GmrS-GmrD, in contrast, has evolved to cleave glucosyl-hmC-containing DNA, including all three hypermodifications discussed above (Bair and Black, 2007; Bair et al., 2007). To

overcome GmrS-GmrD, T-even phages encode an inhibitor, IPI* (Bair and Black, 2007; Bair et al., 2007). Interestingly, the T-even phage RB69 carries arabinosyl-hmC instead of glucosyl-hmC (Thomas et al., 2018), suggesting that differential glycosylation programs may present a strategy to evade type IV RM systems that target a specific DNA modification.

Diverse DNA modifications are present in viruses beyond the T-even family of coliphages (Weigele and Raleigh, 2016). For example, the *Queuovirinae* subfamily of *Siphoviridae* carry hypermodified deazaguanosine nucleobases which provide broad resistance to cleavage by RM systems (Hutinet et al., 2019; Olsen et al., 2023). The *Synechococcus* phage S-2L contains 2-aminoadenine (Khudyakov et al., 1978; Kirnos et al., 1977). *Bacillus* phages PBS1/PBS2 carry uracil (Takahashi and Marmur, 1963) and phages SP8 and SPO1 carry 5-hydroxymethyluracil (Weigele and Raleigh, 2016; Wilhelm and Ruger, 1992), replacing 100% of the thymine nucleobases in their respective viral genomes. DNA modification thus represents a widespread anti-restriction strategy across different families of phages.

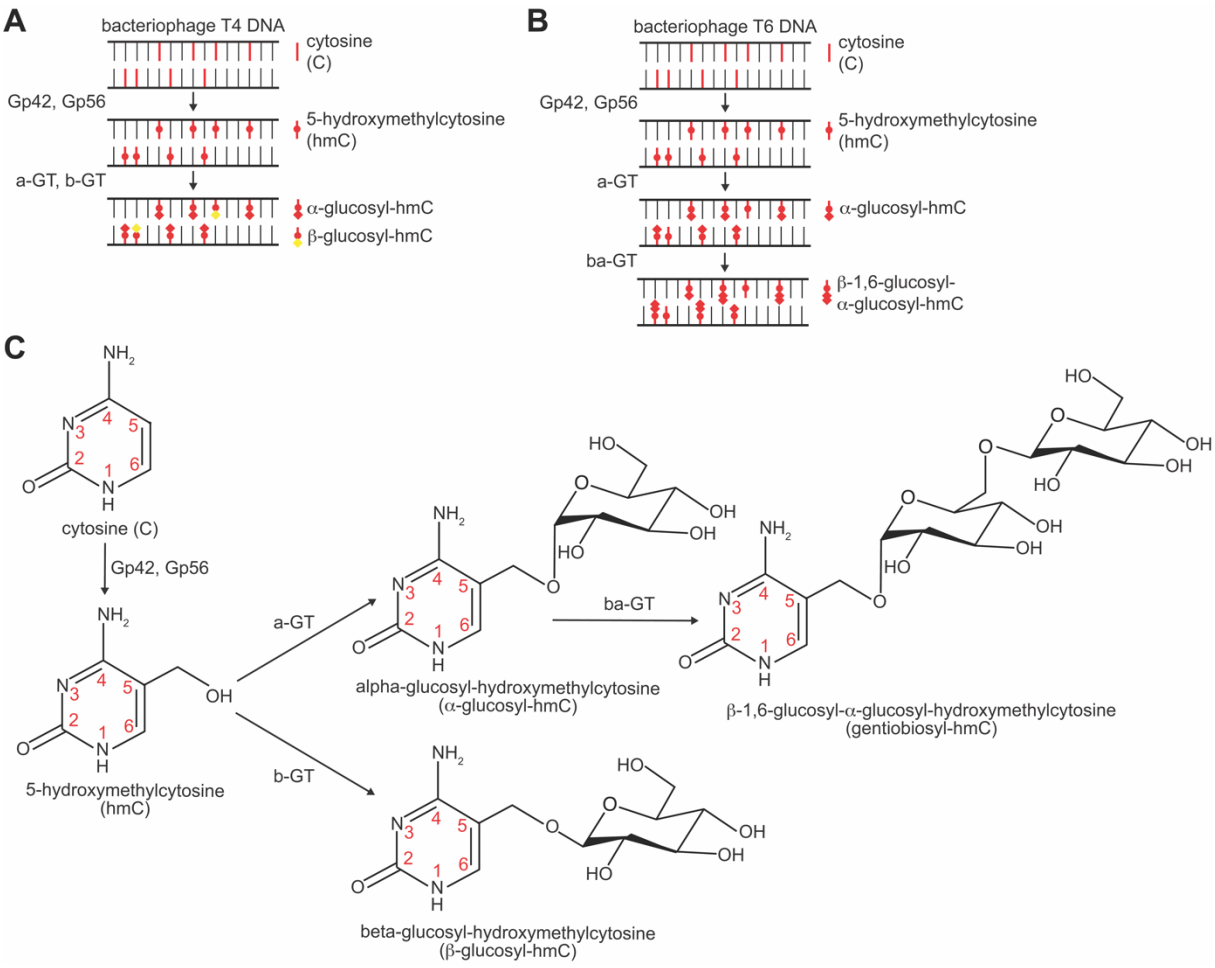


Figure 1.15 Glucosylation of 5-hydroxymethylcytosine nucleobases in T-even phage genomes

(**A-B**) Schematic of the cytosine modification pathway in phages T4 and T6. Phage enzymes: Gp42, dCMP hydroxymethylase; Gp56, dCTPase; a-GT, alpha-glucosyltransferase; b-GT, beta-glucosyltransferase, ba-GT, beta-alpha glucosyltransferase. (**C**) Schematic of the cytosine modification pathway in phages T4 and T6, showing the chemical structures of the nucleobases.

1.15 DNA repair mechanisms in bacteria

DNA damage can be broadly classified into three main categories in bacteria: (i) DNA double-strand breaks (DBSs), (ii) single-strand gaps in DNA, and (ii) nucleotide and single-strand damage. DSBs are repaired in bacteria through homologous recombination (HR), non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ) (Chayot et al., 2010; Cromie et al., 2001; Kowalczykowski et al., 1994; Shuman and Glickman, 2007; Wigley, 2013). HR reconstitutes the original DNA sequence following recombination of the broken DNA with an intact homologous DNA template. This process is mediated by proteins called recombinases. The principal bacterial recombinase is RecA, a homolog of eukaryotic Rad51 (Kowalczykowski, 2000; Lusetti and Cox, 2002; Ogawa et al., 1993; Sung, 1994). In contrast to HR, NHEJ and MMEJ forgo the repair template and ligate broken DNA ends together, either at the site of the break or away from the break following exonucleolytic end resection, which results in a deletion (Chayot et al., 2010; Shuman and Glickman, 2007; Wigley, 2013). Single-strand gaps are repaired primarily through the gap repair pathway, involving RecJ and RecFOR, which recruit RecA to then execute HR (Cox et al., 2023). Nucleotide damage and damage to the sugar-phosphate backbone is repaired through base excision repair (BER), nucleotide excision repair (NER) and mismatch repair pathways (Wozniak and Simmons, 2022). BER uses DNA glycosylases and AP endonucleases to remove damaged and unwanted nucleobases in the DNA backbone (Baute and Depicker, 2008), while NER removes bulky, DNA duplex-distorting nucleotide or strand damage

such as cyclobutane pyrimidine dimers (Sancar and Reardon, 2004). The mismatch repair pathway corrects mismatched DNA bases during replication when DNA polymerase proofreading fails to correct an error (Kunkel and Erie, 2005). For this thesis, I will focus primarily on prokaryotic DSB repair via HR, mediated by bacterial and phage recombinases, and on nucleobase removal via base excision, mediated by DNA glycosylases.

1.16 Degradation and repair of DNA double-strand breaks

Cleavage of phage DNA by Cas9 generates a DSB in the phage genome (Gasiunas et al., 2012; Jinek et al., 2012; Sternberg et al., 2015). How these DSBs are processed in vivo during infection is not known. One expectation is that host nucleases load at the Cas9-induced DSB and further degrade the phage genome, but this degradation has not been directly observed and the nucleases responsible are unknown (Cui and Bikard, 2016). The enzyme RecBCD is the major complex that processes duplex DNA ends in Gram-negative bacteria such as *E. coli* (Dillingham and Kowalczykowski, 2008; Smith, 2012). RecBCD primarily functions as a recombination-repair complex which uses HR to repair stalled or collapsed replication forks and DSBs induced by DNA damaging agents (Dillingham and Kowalczykowski, 2008; Kuzminov, 1995, 1999; Smith, 2012). Somewhat paradoxically, RecBCD also functions as a DNA degradation complex that destroys host and foreign linear DNA that would otherwise be detrimental to the cell (Behme et al., 1976; Miranda and Kuzminov, 2003; Simmon and Lederberg, 1972). The RecBCD complex is made up

of three proteins: RecB which contains an N-terminal 3'-5' helicase and a C-terminal nuclease, RecD which is a 5'-3' helicase, and RecC which recognizes the DNA sequence *chi* (crossover hotspot instigator) (Bianco and Kowalczykowski, 1997; Henderson and Weil, 1975; Lam et al., 1974) to stimulate RecA-mediated homologous recombination (Dillingham and Kowalczykowski, 2008; Smith, 2012). RecBCD binds with high affinity to the blunt or nearly blunt free end of linear duplex DNA or to free ends of a DSB (Taylor and Smith, 1985) (**Fig. 1.16**). The enzyme moves along DNA, with high speed and processivity, using ATP hydrolysis to fuel translocation and unwinding (Dillingham et al., 2003; Taylor and Smith, 1985, 2003). During translocation, the RecB C-terminal nuclease domain introduces single-strand nicks into both unwound DNA strands, generating ssDNA degradation products (Dixon and Kowalczykowski, 1993; Sun et al., 2006; Wang et al., 2000). RecBCD continues translocation and degradation of DNA in this manner until it encounters a *chi* sequence in the proper orientation (Bianco and Kowalczykowski, 1997). In *E. coli*, *chi* is the octameric single-stranded DNA sequence 5'-GCTGGTGG-3' (Bianco and Kowalczykowski, 1997) and is statistically overrepresented in the *E. coli* genome (1 per 4.5 kb) (Touzain et al., 2011). When RecBCD encounters a *chi* site, the enzyme switches its mode of action from degradation to repair (Anderson and Kowalczykowski, 1997a; Dixon and Kowalczykowski, 1991, 1993, 1995; Spies et al., 2003). The *chi* site is recognized from the 3' end by a *chi* scanning tunnel in RecC (Amundsen et al., 2016; Handa et al., 2012; Singleton et al., 2004). Upon *chi* recognition, the RecB nuclease stops degradation of the 3' ssDNA tail. The enzyme

continues to translocate beyond *chi*, degrading only the strand with the 5' ssDNA tail (Anderson and Kowalczykowski, 1997a; Dixon and Kowalczykowski, 1991, 1993, 1995; Ponticelli et al., 1985; Spies et al., 2003). This creates a looped 3' overhang on the top strand containing *chi*. RecB now loads RecA onto the looped ssDNA overhang (Anderson and Kowalczykowski, 1997b; Arnold and Kowalczykowski, 2000; Spies et al., 2005). RecA monomers loaded by RecBCD result in RecA filament formation in the 5'-3' direction on the overhang strand up to *chi* (Galletto et al., 2006). RecBCD eventually dissociates, and the RecA-ssDNA complex initiates homologous recombination with a sister DNA template (Dixon and Kowalczykowski, 1991; Roman and Kowalczykowski, 1989). Recombination proceeds through strand invasion of homologous DNA by the RecA nucleoprotein filament followed by exchange of DNA strands via formation and resolution of two Holliday junctions (Kuzminov, 1999; Lenhart et al., 2012; West, 1994) (**Fig. 1.17**). In Gram-positive bacteria such as *B. subtilis* and *S. pyogenes*, AddAB serves as the functional homolog of RecBCD and similarly performs the roles of degradation and RecA-mediated repair in a *chi*-dependent manner (Chedin et al., 2006; Chedin et al., 1998; Wigley, 2013).

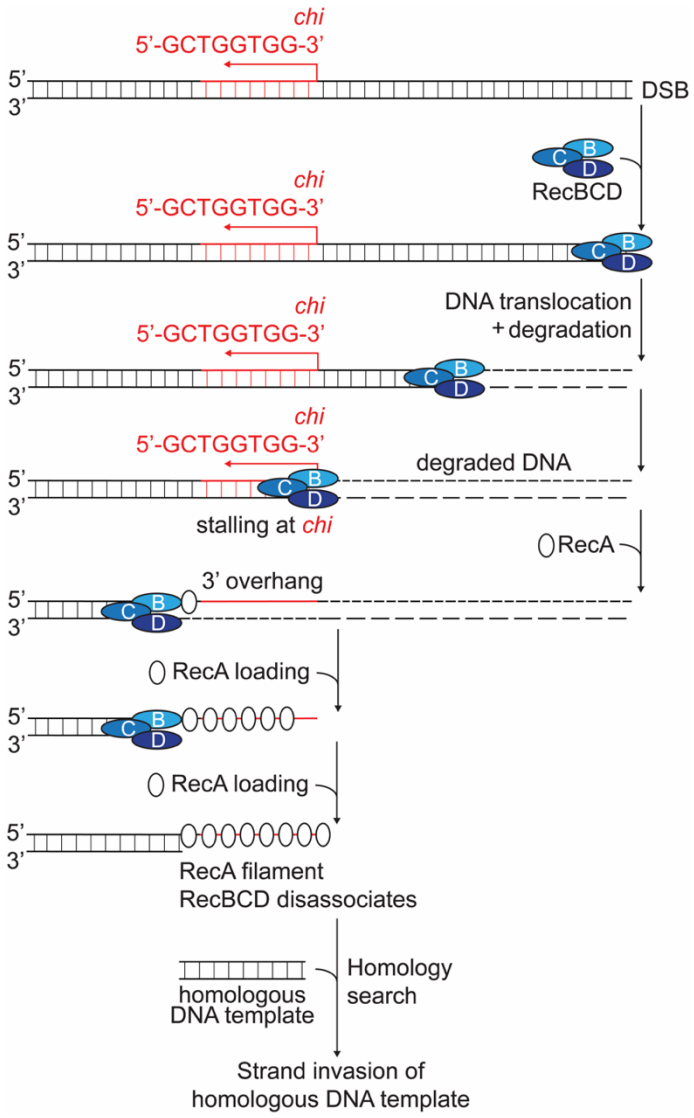


Figure 1.16 Degradation and repair of DNA double-strand breaks by RecBCD

RecBCD binds with high affinity to a free dsDNA end. The enzyme translocates along DNA, unwinding the DNA duplex and degrading both top and bottom strands using the C-terminal nuclease tail of RecB. The faster translocation speed of the RecD helicase relative to the RecB helicase results in longer ssDNA degradation products being generated from the bottom strand. Upon recognition of a *chi* sequence by RecC, the enzyme stalls and the RecB nuclease stops degradation of the top strand. The complex continues to translocate beyond *chi* while only degrading the bottom strand as shown. This generates a 3' overhang on the top strand. RecB loads RecA onto the 3' overhang, and as more RecA is added in the 5'-3' direction, a RecA-ssDNA filament is formed. RecBCD ultimately dissociates and the RecA-ssDNA filament mediates strand invasion of a homologous DNA template. Figure adapted from Dillingham and Kowalczykowski (2008) and reproduced with permission from American Society for Microbiology.

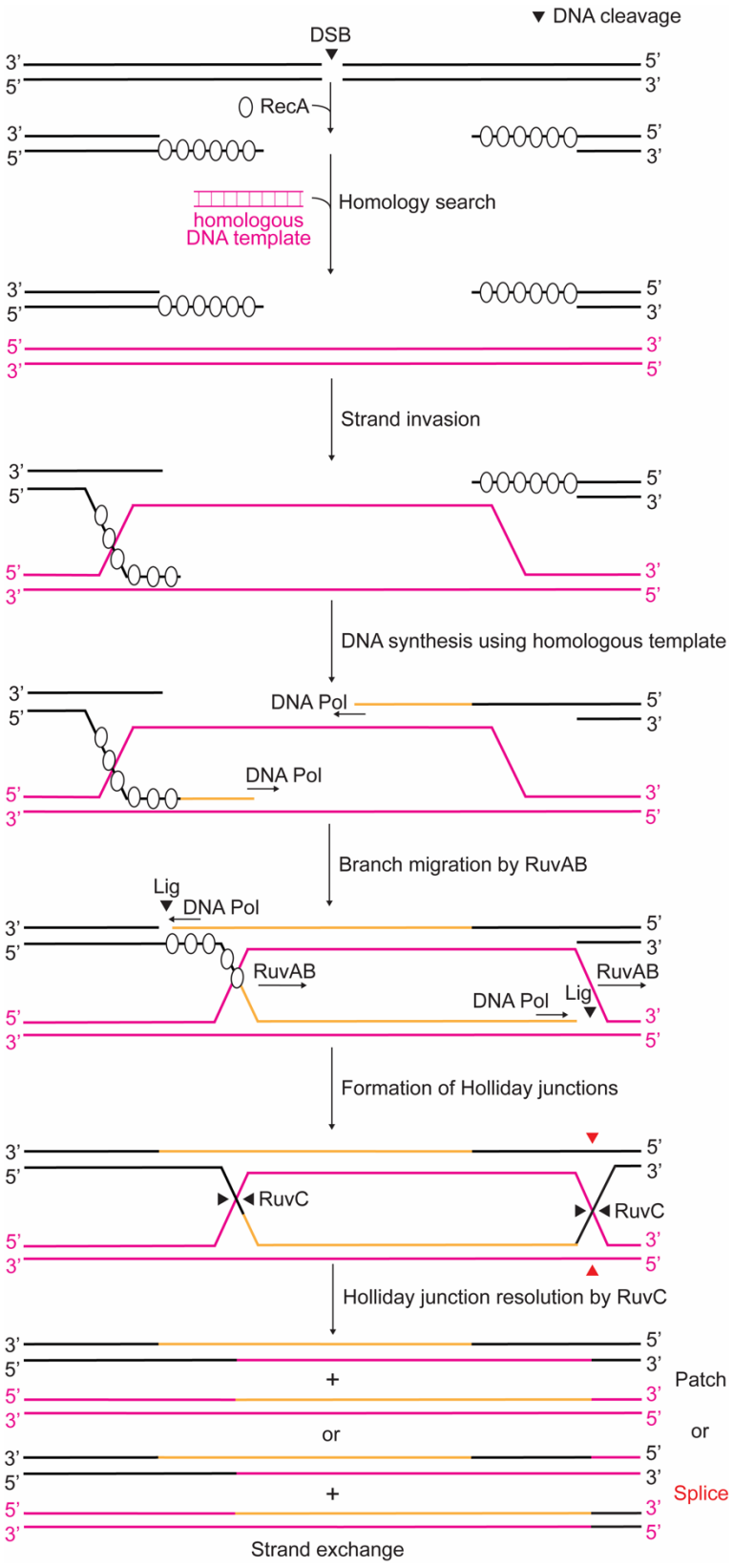

▼ DNA cleavage
DSB
3'
5'
5'
3'
RecA
homologous
DNA template
Homology search
Strand invasion
DNA synthesis using homologous template
DNA Pol
Branch migration by RuvAB
Lig
RuvAB
Formation of Holliday junctions
RuvC
Holliday junction resolution by RuvC
+
or
Patch
Splice
Strand exchange

Figure 1.17 Homologous recombination by RecA through formation of double Holliday junctions

Following RecBCD resection of both ends of a DSB and *chi*-mediated RecA loading on either side of the DSB, a RecA-ssDNA filament mediates strand invasion of a homologous DNA template to initiate repair through formation of a displacement loop (D loop). Formation of the D loop allows DNA strands from the homologous donor DNA to serve as templates through strand exchange to initiate DNA synthesis. The RuvAB complex mediates migration of the D loop as DNA synthesis proceeds, extending the degree of strand exchange. DNA polymerase elongates DNA strands and DNA ends are ligated, yielding parental and donor strands that form crossover structures called Holliday junctions. Holliday junctions are resolved by the RuvC resolvase, which can cleave daughter or parental strands to give strand exchanged recombination productions. Figure adapted from Lenhart et al. (2012) and reproduced with permission from American Society for Microbiology.

1.17 The Red recombination system of phage λ

Phages often encode their own recombination systems that compete with host machinery such as RecBCD to act on DSBs. An example is the phage λ recombination system Red (Echolas and Gingery, 1968; Signer and Weil, 1968) (**Fig. 1.18-A**). The Red system comprises the operon of genes *gam*, *bet* and *exo* which encode the Gam, Beta and Exo proteins, respectively (Murphy, 2016). Gam prevents degradation of linear phage DNA or free DNA ends by inhibiting RecBCD (Court et al., 2007; Karu et al., 1975; Murphy, 1991, 2007), allowing Exo-Beta recombination to act on free DNA ends and DSBs. Exo is a highly processive 5'-3' exonuclease that binds to a free DNA end and extensively degrades the 5' strand to generate a long 3' ssDNA overhang (Little, 1967; Little et al., 1967; Sriprakash et al., 1975). Beta binds to the overhang DNA and coats the ssDNA to form a nucleoprotein filament (Passy et

al., 1999). Unlike RecA which mediates strand invasion, Beta mediates recombination through single-strand annealing of the Beta-ssDNA filament to a complementary ssDNA partner (Karakousis et al., 1998; Kmiec and Holloman, 1981; Muniyappa and Radding, 1986). The Beta-ssDNA complex is thought to pair with the lagging strand of a replication fork (Mosberg et al., 2010; Poteete, 2008). Red has an important role in the phage λ lytic cycle: Exo and Beta are thought to facilitate phage concatemer formation after theta replication, while Gam prevents RecBCD degradation of the free DNA ends of linear concatemers generated during rolling-circle replication (Smith, 1983; Smith, 2012) (**Fig. 1.18-B**).

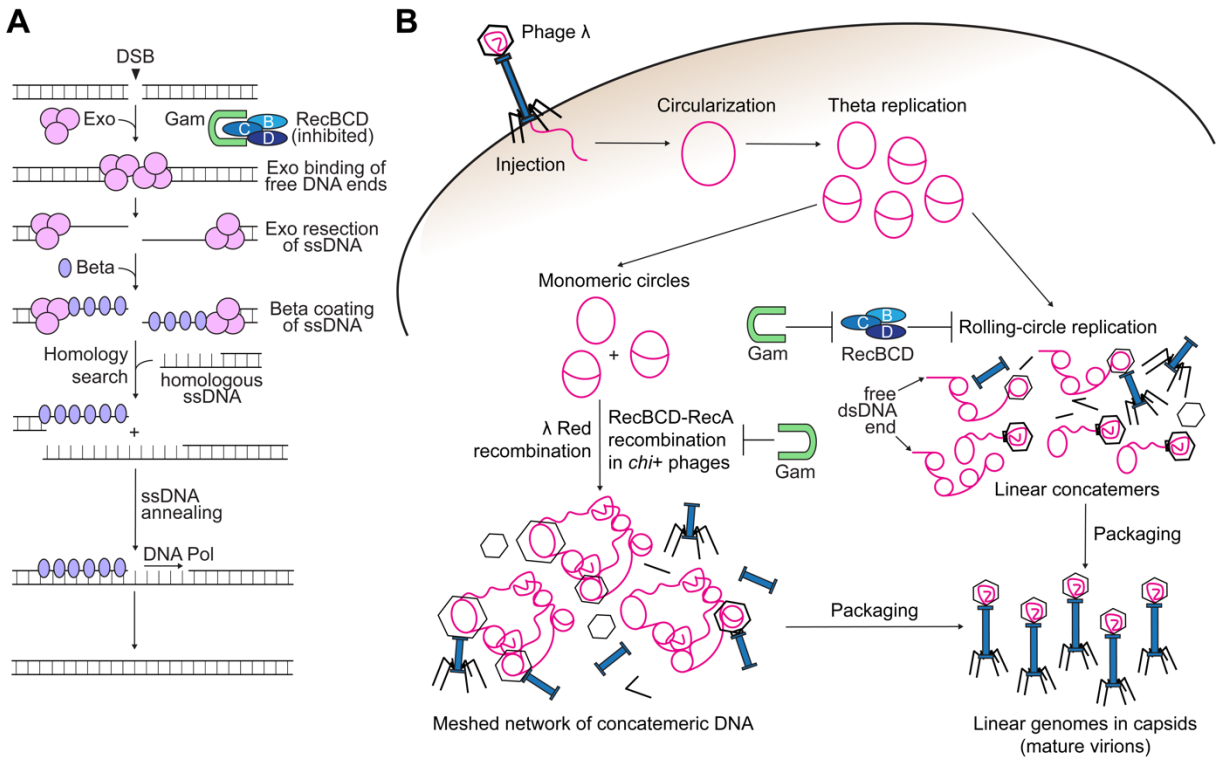


Figure 1.18 λ Red recombination

(**A**) λ Red recombination involves the proteins Gam, Exo and Beta. Gam prevents degradation of free DNA ends by inhibiting RecBCD. Exo binds to free DNA ends and processively degrades single strands to generate long 3' ssDNA overhangs. Beta coats the ssDNA overhangs and mediates recombination through a single-strand annealing mechanism. (**B**) λ Red plays a role in the viral lytic cycle. Upon injection, the linear phage genome circularizes and undergoes theta replication to produce monomeric circular genomes. Monomers can form concatemers necessary for viral packaging through λ Red recombination or RecBCD-RecA recombination if the phage contains a viable *chi* site. In late infection, the phage undergoes unidirectional rolling-circle replication to form linear concatemeric DNA that is packaged into capsids. Gam inhibits RecBCD, which would otherwise degrade the linear concatemers by loading onto their free DNA ends. Panel **B** adapted from Smith (2012) and reproduced with permission from American Society for Microbiology; Smith (2012) figure based on Smith (1983) [copyright 1983, Cold Spring Harbor Laboratory Press].

1.18 End joining pathways to repair DNA double-strand breaks

End joining pathways forgo homologous repair templates and directly ligate together blunt or resected DNA ends (Wigley, 2013). Non-homologous end joining (NHEJ) is thought to dominate HR in quiescent cells where there is lack of sister DNA templates due to the absence of DNA replication (Shuman and Glickman, 2007). Much of our knowledge of bacterial NHEJ comes from studies in mycobacteria (Shuman and Glickman, 2007). The DNA-end binding protein Ku is the central player in NHEJ (Weller et al., 2002). Ku finds the DNA ends generated during DSB formation and brings them together for repair. A specialized DNA ligase, LigD, ligates the broken DNA ends together (Gong et al., 2005; Shuman and Glickman, 2007) (**Fig. 1.19-A**). The repair can be error-free, but is often mutagenic, involving single-

nucleotide insertions or short deletions (Gong et al., 2005). A less studied end joining mechanism in bacteria is microhomology-mediated end joining (MMEJ) and the proteins involved are not clearly defined, although certain bacterial ligases have been implicated in *E. coli* (Chayot et al., 2010). In MMEJ, the free DNA ends are resected to give single-strand overhangs which expose short 5-25 bp complementary sequences, referred to as microhomology (McVey and Lee, 2008), on both ends of a DSB, now exposed as ssDNA due to end resection. The microhomology sequences are paired through complementary base pairing, resulting in a deletion of the sequence in between the microhomology sequences (Seol et al., 2018) (**Fig. 1.19-B**).

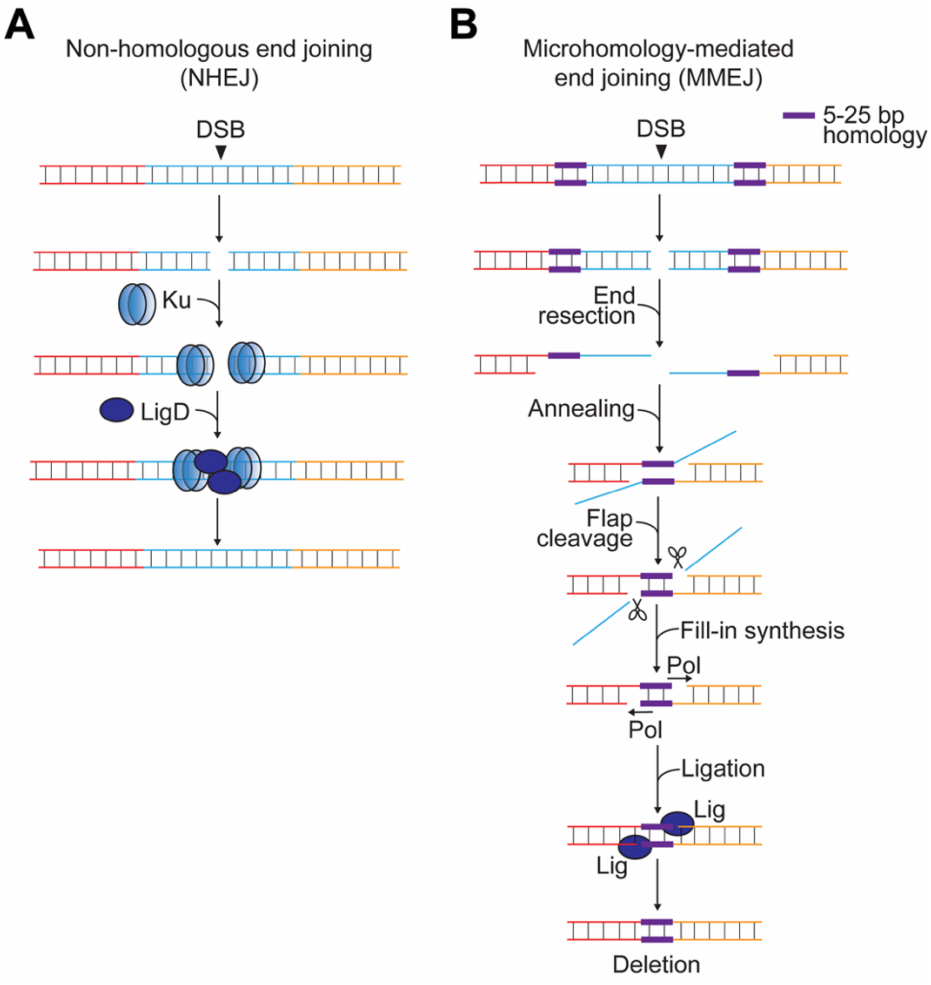


Figure 1.19 End joining repair pathways in bacteria

(**A**) Non-homologous end joining (NHEJ) uses the DNA-end binding protein Ku to locate broken DNA ends and bring them together to be ligated by the specialized DNA ligase LigD. Panel **A** adapted from Shuman and Glickman (2007) and reproduced with permission from Springer Nature. (**B**) Microhomology-mediated end joining (MMEJ) involves end resection at either end of a DSB, which generates ssDNA overhangs and exposes 5-25 bp complementary sequences on the ssDNAs on opposite ends of the original DSB. The sequences are annealed together, and the protruding ssDNA flaps are cleaved. Gaps are filled in and ligated, and the result is a deletion of the DNA sequence between the two microhomology sequences. Panel **B** adapted from Seol et al. (2018) and reproduced with permission from Elsevier.

1.19 Base excision repair

Base excision repair (BER) is a DNA repair pathway that removes and replaces damaged and non-canonical DNA nucleobases (Baute and Depicker, 2008; Krokan and Bjoras, 2013; Thompson and Cortez, 2020; Wozniak and Simmons, 2022). BER constitutes the following series of concerted steps: recognition and excision of the target nucleobase, single strand cleavage of the DNA sugar-phosphate backbone, fill-in DNA synthesis and cleavage of displaced DNA followed by ligation of the ssDNA gap (Thompson and Cortez, 2020) (**Fig. 1.20-A**). BER is initiated by enzymes called DNA glycosylases. These recognize and excise a target nucleobase from DNA backbones to generate an abasic site (**Fig. 1.20-B**). The most recognized of these enzymes are the uracil DNA glycosylases, which excise uracil from DNA to generate an abasic site that can be repaired via BER (Pearl, 2000; Schormann et al., 2014). DNA glycosylases excise their target bases by hydrolyzing the N-glycosidic bond

between the nucleobase and the sugar-phosphate backbone. Monofunctional DNA glycosylases facilitate base removal while leaving the sugar-phosphate backbone intact. In contrast, bifunctional glycosylases possess a lyase activity that catalyzes strand cleavage after base removal (Jacobs and Schar, 2012; Sun et al., 1995). Specifically, they cleave the DNA sugar-phosphate backbone immediately 3' of the glycosylase-generated abasic site, in a reaction known as beta-elimination (Bailly and Verly, 1987; Talpaert-Borle, 1987; Thompson and Cortez, 2020) (**Fig. 1.20-C**). In the absence of lyase activity, monofunctional glycosylases rely on a separate enzyme known as an apurinic/apyrimidinic (AP) endonuclease to cleave the DNA backbone following base excision (**Fig. 1.20-A**). AP endonucleases cleave the sugar-phosphate backbone 5' of the abasic site (Thompson and Cortez, 2020). Following strand cleavage, two downstream repair subpathways can occur: short-patch or long-path repair (Fortini and Dogliotti, 2007; Petermann et al., 2003; Sukhanova et al., 2005; Thompson and Cortez, 2020). Short-patch repair is the predominant pathway in bacteria and requires the RecJ exonuclease in *E. coli* to excise the 5' terminal deoxyribose-phosphate generated following AP endonuclease-mediated strand cleavage (Dianov and Lindahl, 1994; Dianov et al., 1994). A DNA polymerase, typically Pol I in *E. coli*, fills in the single nucleotide gap with the correct nucleotide and a ligase seals the nick to complete repair (Dianov and Lindahl, 1994). Long-patch repair occurs when multiple nucleobases within a single stretch of DNA require replacement by the BER pathway (Fortini and Dogliotti, 2007; Petermann et al., 2003; Sukhanova et al., 2005; Thompson and Cortez, 2020). Following strand cleavage, a

DNA polymerase performs strand displacement synthesis to generate a 5' flap, which is subsequently cleaved by a flap endonuclease. As before, DNA ligase seals the nick.

The deamination of cytosine is one of the major forms of DNA damage in cells, resulting in the conversion of cytosine to uracil (Lindahl, 1993; Thompson and Cortez, 2020). This causes a U:G mismatch in the DNA, which distorts the DNA helix and can result in a C:G→T:A transition mutation if left unrepaired prior to replication (Pearl, 2000). BER of uracil nucleobases is initiated by uracil DNA glycosylases, making them amongst the most widespread DNA repair enzymes across the domains of life. The uracil DNA glycosylases comprise a superfamily of conserved enzymes that specialize in the excision of uracil from ssDNA and dsDNA backbones. There are six different families of uracil DNA glycosylases, classified according to their substrates (Pearl, 2000; Schormann et al., 2014). While they vary in substrate preference, all known members of the uracil DNA glycosylase superfamily are monofunctional glycosylases that lack a secondary AP lyase activity (Schormann et al., 2014). These enzymes bind DNA non-specifically and perform target search through a combination of DNA hopping and one-dimensional short-range sliding (Hedglin and O'Brien, 2010; Porecha and Stivers, 2008). Like most glycosylases, these enzymes use a nucleotide-flipping mechanism for recognition and excision of their target nucleobase (Slupphaug et al., 1996; Stivers, 2004). Upon recognition of a uracil nucleobase, recognized as a extrahelical distortion of the DNA duplex by the

human uracil DNA glycosylase hUNG (a family I glycosylase), the enzyme changes from an open to a closed confirmation (Schormann et al., 2014), resulting in flipping of the deoxyuridine nucleotide in the binding pocket of the glycosylase (Mol et al., 1995; Slupphaug et al., 1996). This is followed by hydrolysis of the N-glycosidic bond between the base and the sugar, resulting in release of the free uracil base (Schormann et al., 2014). In hUNG, the catalytic aspartate residue in the substrate pocket activates a water molecule through proton abstraction to generate a hydroxide ion that performs a nucleophilic attack on the N-glycosidic bond of deoxyuridine (Slupphaug et al., 1996). In contrast, family II mismatch-specific uracil DNA glycosylases, such as *E. coli* MUG, use water, which is weakly nucleophilic, rather than a hydroxide ion to attack the N-glycosidic bond, resulting in slower enzyme chemistry (Barrett et al., 1998).

Abasic sites are highly reactive both in vitro and in vivo (Thompson and Cortez, 2020). They exist in an equilibrium between a closed-ring furanose (99%) and an open-ring aldehyde (1%) (Wilde et al., 1989), where the latter is highly reactive and its 3' phosphodiester bond is prone to cleavage (Lhomme et al., 1999; Thompson and Cortez, 2020) (**Fig. 1.20-C**). Abasic sites generated in vitro can result in strand cleavage through beta-elimination (Bailly and Verly, 1987; Talpaert-Borle, 1987; Thompson and Cortez, 2020). Treatment of the abasic site-containing DNA with heat or a base such as sodium hydroxide (NaOH) promotes strand cleavage of the sugar-phosphate backbone 3' of the abasic site by beta-elimination (Miyazono et al., 2014;

Talpaert-Borle, 1987) (**Fig. 1.20-C**). Abasic sites in vivo are highly deleterious to genomic stability because of their reactive nature (Thompson and Cortez, 2020). They can result in strand breaks through beta-elimination, block the progression of DNA and RNA polymerases or react with DNA or proteins to form stable or transient covalent linkages (Thompson and Cortez, 2020) (**Fig. 1.20-C**). Stalling of RNA polymerases prevents transcription of DNA and hence gene expression, while blocked DNA polymerases prevent DNA replication. The open-ring aldehyde form of an abasic site can also react with other nucleobases to form DNA intra- and interstrand crosslinks (Dutta et al., 2007; Price et al., 2014; Sczepanski et al., 2008; Thompson and Cortez, 2020). It can also react with proteins such as DNA polymerases to form stable DNA-protein crosslinks (DPCs) (Quinones and Demple, 2016) as well as transient Schiff-base intermediates with proteins such as polymerases, glycosylases, and other DNA repair proteins (Thompson and Cortez, 2020). Given the wide variety of DNA lesions generated via abasic sites, the repair of these sites is imperative to both genome integrity and cell viability and takes place through the BER and NER pathways.

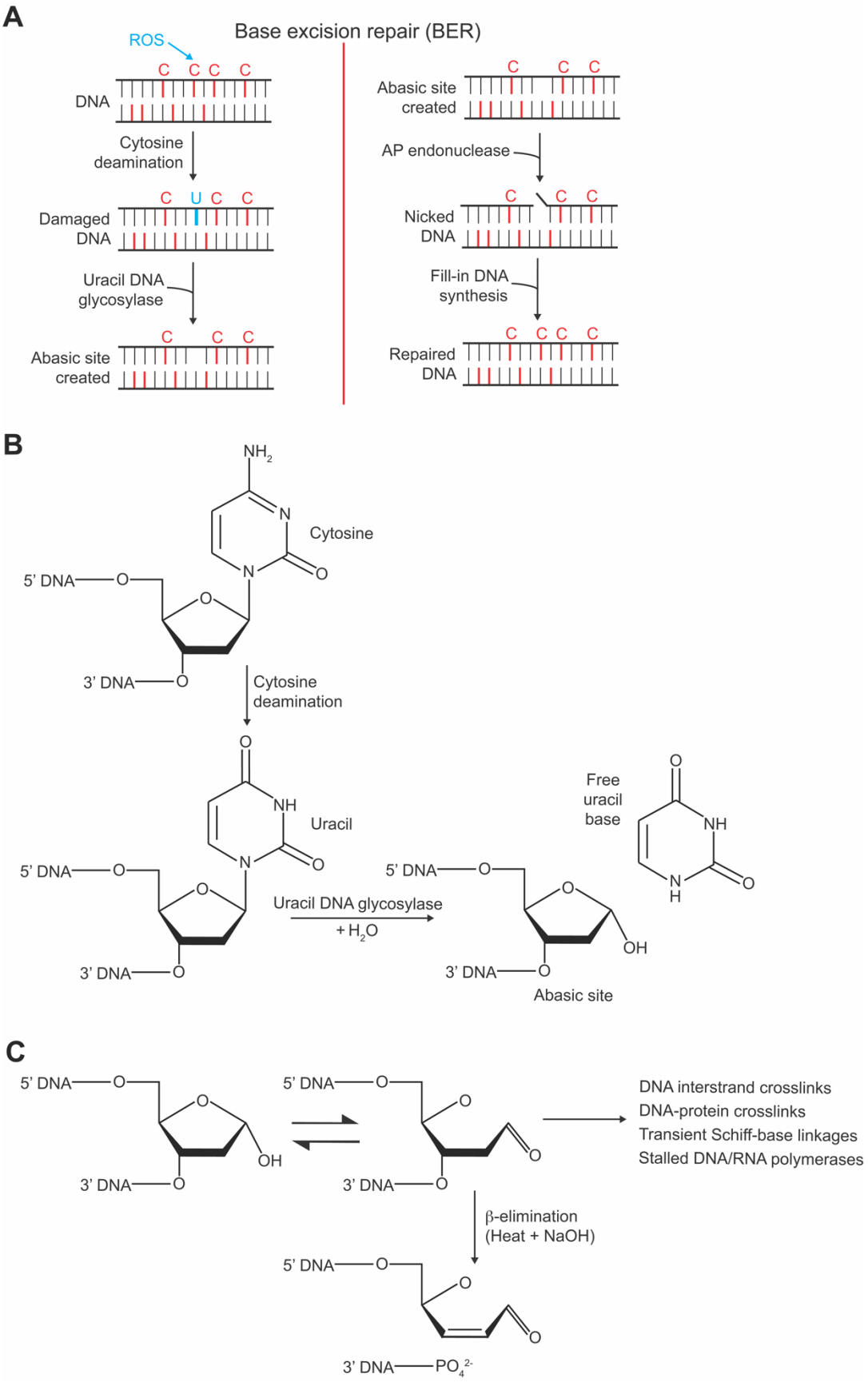

A
Base excision repair (BER)
ROS
DNA
Cytosine deamination
Damaged DNA
Uracil DNA glycosylase
Abasic site created
Abasic site created
AP endonuclease
Nicked DNA
Fill-in DNA synthesis
Repaired DNA
B
Cytosine
5' DNA
3' DNA
Cytosine deamination
Uracil
Uracil DNA glycosylase
+ H2O
Free uracil base
Abasic site
C
DNA interstrand crosslinks
DNA-protein crosslinks
Transient Schiff-base linkages
Stalled DNA/RNA polymerases
β-elimination
(Heat + NaOH)

Figure 1.20 Base excision repair

(**A**) Schematic of the base excision repair pathway involving removal of uracil by uracil DNA glycosylase. Reactive oxygen species (ROS) can deaminate cytosine to uracil. Uracil in DNA is recognized by uracil DNA glycosylase, which excises the base while leaving the sugar-phosphate backbone intact to generate an abasic site. The DNA backbone is cleaved at the 5' end of the abasic site by an apurinic/apyrimidinic (AP) endonuclease. In short-patch repair (shown here), a DNA polymerase fills the single nucleotide gap and a DNA ligase seals the nick. (**B**) Chemical structures of cytosine, uracil and an abasic site generated by uracil DNA glycosylase. (**C**) Consequences of abasic sites in vivo. Abasic sites are highly reactive and can form a variety of deleterious covalent linkages, including interstrand DNA crosslinks, stable DNA-protein crosslinks, and transient DNA-protein linkages, can stall DNA/RNA polymerases and can result in strand breakage through beta-elimination. Panels **B** and **C** adapted from Thompson and Cortez (2020) and reproduced with permission from Elsevier.

1.20 DNA repair in the bacteria-phage arms race

In summary, bacteria possess a vast arsenal of immune defenses that target the DNA of invading phages (Georjon and Bernheim, 2023) (**Fig. 1.7** and **Fig. 1.8**). These systems induce DNA damage in phage genomes, including DSBs (**Fig. 1.8**), raising the possibility of phages exploiting a variety of DNA repair mechanisms to counter the damage inflicted by bacterial anti-phage systems. In the next two chapters, I will describe experimental results which demonstrate that bacteriophages exploit DNA recombination systems to repair CRISPR-Cas cleaved viral DNA and evade CRISPR-Cas targeting (Hossain et al., 2021). These results suggest that DNA repair is yet another viral evasion strategy to go along with the numerous counter-immune strategies described earlier in this chapter (**Fig. 1.12**). In the later chapters of this thesis, I will describe how bacteria have co-opted a specialized DNA glycosylase,

likely evolved from the uracil DNA glycosylase superfamily, to execute antiviral defense against phages that carry alpha-glucosyl-hydroxymethylcytosine DNA modifications. The excision of hypermodified nucleobases represents a novel mechanism of DNA targeting by a bacterial immune system, one that overcomes the viral anti-restriction strategy of modifying nucleobases (**Fig. 1.15**) to prevent cleavage by CRISPR-Cas and restriction endonucleases (Bryson et al., 2015; Kruger and Bickle, 1983; Liu et al., 2020b; Vlot et al., 2018). More importantly, this discovery represents a unique repurposing of a DNA repair protein for defense functionality. Overall, my work opens the door for delving further into the co-option and roles of DNA repair modules in prokaryotic host-virus conflicts.

CHAPTER 2. BACTERIOPHAGE RECOMBINATION MEDIATES EVASION OF CRISPR-CAS TARGETING

2.1 Background

Clustered regularly interspaced short palindromic repeat (CRISPR) loci and CRISPR-associated (*cas*) genes protect bacteria and archaea against foreign genetic elements such as viruses (Barrangou et al., 2007) and plasmids (Marraffini and Sontheimer, 2008). Upon infection, short invader sequences, known as spacers, are inserted in between the repeats of the CRISPR locus (Barrangou et al., 2007). These are subsequently transcribed and processed to generate short CRISPR RNAs (crRNAs) (Brouns et al., 2008; Hale et al., 2008; Tang et al., 2005) that are used as guides by Cas complexes to recognize and destroy complementary protospacer sequences within the nucleic acids of the invading virus or plasmid (Garneau et al., 2010; Gasiunas et al., 2012; Hale et al., 2009; Jinek et al., 2012; Jore et al., 2011).

CRISPR-Cas systems can be classified into six different types depending on their *cas* gene content (Makarova et al., 2020). Types I and II are the most common DNA-cleaving systems and have two target requirements for activity: a protospacer-adjacent motif (PAM) and a seed sequence within the protospacer. In the commonly studied type II-A system of the Gram-positive bacterium *S. pyogenes*, the PAM is a 5'-NGG-3' sequence immediately downstream of the protospacer and the seed sequence is located in the 6-8 nucleotides that precede the PAM (Bikard et al., 2012; Jinek et al., 2013). Successful recognition of a target by the Cas9 RNA-guided

nuclease leads to the introduction of a double-strand break (DSB) in the protospacer DNA three nucleotides upstream of the PAM (Garneau et al., 2010)}(Jinek et al., 2013). In the *E. coli* type I-E system, the RNA-guided Cascade complex recognizes targets with a 5'-AWG-3' PAM (Westra et al., 2012) upstream of the protospacer and a seed sequence in the 8 nucleotides immediately downstream of the PAM (Semenova et al., 2011). Upon target recognition, Cascade recruits the ssDNA nuclease Cas3 (Westra et al., 2012), which first degrades both complementary and non-complementary DNA strand (Mulepati and Bailey, 2013; Sinkunas et al., 2013) and then unwinds and further degrades the non-complementary strand (Mulepati and Bailey, 2013; Redding et al., 2015). These activities not only cut, but also further degrade the target DNA; however, the extent to which Cas3 or other host nucleases are involved in the destruction of the invader's genome is not known. In both systems, mutations in either the PAM or seed sequences lead to evasion of CRISPR immunity by the invader (Deveau et al., 2008; Nussenzweig et al., 2019; Semenova et al., 2011). A recent study investigating type II-A immunity against T4 phage in *E. coli* showed that viral escape mutations accumulated during the course of infection (Tao et al., 2018), suggesting that target mutations are unlikely pre-existing but rather introduced *de novo* after Cas9 cleavage. However, how these mutations are generated and whether DNA repair and recombination play a role was not known at the outset of my thesis work. The role of viral DNA repair in the evasion of CRISPR-Cas targeting and in the generation of viral escape mutations is thus the focus of the first study in my thesis.

2.2 Using phage λ to probe the effects of phage recombination on CRISPR-Cas targeting

In this chapter of my PhD thesis, I investigate whether and how DNA recombination systems encoded by bacteriophages mediate evasion of CRISPR-Cas immunity in bacteria (Hossain et al., 2021). In the next chapter, I look at whether recombinational repair can result in *de novo* target mutations that facilitate viral escape. This work was a collaborative effort with Dr. Joshua Modell, who was a postdoctoral fellow transitioning to a tenure-track assistant professor position during my first year in the Marraffini lab. To investigate the impact of recombination systems on the outcome of CRISPR targeting, we decided to study the effects of the best characterized RNA-guided DNA nuclease, Cas9 (Jiang and Doudna, 2017) on the most studied bacterial virus, the λ phage (Wegrzyn et al., 2012). Specifically, we probed how phage λ escapes cleavage by type II-A and type I-E CRISPR-Cas systems that we genetically programmed in *E. coli* to target phage λ. Both the host and its invader possess pathways that can hypothetically repair the DSBs generated by Cas nucleases through homologous recombination (Wright et al., 2018). First, an exonuclease recognizes the free DNA ends generated at the break and degrades each DNA strand asymmetrically. This introduces a 3' end overhang that is subsequently covered by a ssDNA binding protein that mediates recombination with a homologous DNA template. In *E. coli*, the main homologous recombination pathway is RecABCD, where RecBCD is the exonuclease and RecA is the recombinase (Dillingham and Kowalczykowski, 2008; Kuzminov, 1999). RecBCD

binds to free dsDNA ends at a DSB, then rapidly and processively degrades both strands of DNA until it reaches a *chi* site, beyond which asymmetric degradation of dsDNA by RecBCD generates a 3' overhang (Dillingham and Kowalczykowski, 2008). RecA coats the overhang and mediates recombination via strand invasion into an intact DNA molecule with a homologous sequence. Following branch migration, DNA polymerases fill in the gaps and the Holliday junctions are resolved.

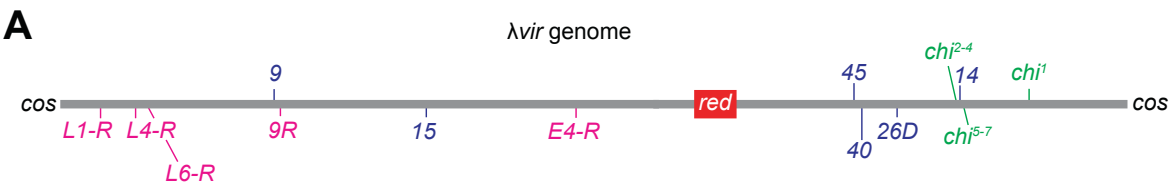
Phage λ harbors the *red* operon, which contains three genes: *gam*, *exo* and *bet*. *exo* encodes the exonuclease (Exo) that asymmetrically degrades free DNA ends to generate 3' overhangs and *bet* produces the recombinase of this system (Beta) that coats the ssDNA for recombination via single-strand annealing of two homologous templates (Mosberg et al., 2010). *gam* encodes a RecBCD inhibitor (Gam) (Kuzminov, 1999) which enables Exo-Beta to drive recombination during phage λ infection (Murphy, 1998). Previous studies found that the Red system is important for phage λ replication (Echolas and Gingery, 1968; Signer and Weil, 1968), to generate long genome concatemers during rolling circle replication (Enquist and Skalka, 1973). Notably, the Red system is not essential and its absence can be rescued by the introduction of an *E. coli chi* site into the λ genome, which mediates recombination via the host's RecABCD system (Henderson and Weil, 1975). Therefore, the widespread distribution of phage recombinases in a large number of both temperate and lytic phage genomes (Lopes et al., 2010) suggests they possess

additional functions that select for their presence in otherwise highly compact phage genomes (Brüssow and Hendrix, 2002).

In this study, we found that many CRISPR-Cas spacers, mediating the targeting of both type I and type II effector complexes (Cascade-Cas3 and Cas9 respectively) against bacteriophage λ, allowed the propagation of large numbers of escape phages with target mutations, and thus provided poor defense. We found that escaper formation required Exo-Beta recombination to generate *de novo* escape mutations. Interestingly, genetically manipulating both the host and the phage to enable robust RecABCD recombination, and hence repair of the CRISPR-cleaved phage DNA through the RecABCD pathway, failed to generate escapers as efficiently as the λ Red recombination pathway. Our results define an additional function for the λ Red system, and likely for other similar phage recombination systems, in driving escape from CRISPR-Cas targeting through the mutagenic repair of DSBs.

2.3 Escape mutations within phage λ Cas9 targets are generated during infection

To study CRISPR-Cas9 targeting of phage λ in *E. coli*, we programmed pCas9 (Jiang et al., 2013), an *E. coli* plasmid that carries the type II-A Cas9 RNA-guided nuclease from *Streptococcus pyogenes* SF370 and its co-factor tracrRNA, with six crRNA guides that target different genomic regions of a virulent mutant of phage λ (λ*vir*) (**Fig. 2.1-AB**).



B

spacer *spcNT*	position	protospacer	PAM
spc9	(+)9187	AGATACCAGCTTCACGCTGGCGTGGATGCC	CGG
spc40	(-)33043	AGGGATGCACCATTCTGAGATGTTTTTATT	TGG
spc45	(+)32965	ACTCGTCAGAATGAATATTATCAAGCAGCA	AGG
spc14	(+)36057	CAGGGGTGTTACCACTACCGCAGGAAAAGG	AGG
spc15	(-)19979	CACCATCAGTTCAAGACGACGCAGCACCTC	CGG
spc26D	(-)34114	TACGGCGCAATTCCGCATCAGTAAGCGCAT	TGG

Figure 2.1 Targeting of phage λ*vir* with different CRISPR-Cas spacers

(**A**) Genomic map of the λ*vir* phage used in this study (roughly to scale) showing the different Cas9 (blue) and Cascade-Cas3 (pink) spacer targets, as well as *chi* sites (in green), used throughout the study. *cos*, cohesive end. The location of the *red* genes is shown as a red box. **(B)** Sequences of the type II-A targets used in this study.

We found that only one of the six spacers above mediated robust immunity against the phage (*spc45*, with an efficiency of plaquing, EOP, of ~10^{-3}; **Fig. 2.2-AB**). The other five spacers either mediated poor (*spc15* and *spc26D*) or extremely weak (*spc9*, *spc40* and *spc14*) immunity. DNA sequencing revealed that phages from all plaques analyzed contained escape mutations that prevent Cas9 cleavage (**Fig. 2.2-C**). Escapers of *spc9*, *spc40* and *spc45* displayed single-nucleotide modifications in either the target seed or PAM sequences. Escapers of *spc14* harbored deletions between short homologous sequences, generated through microhomology-mediated

end joining (MMEJ), a poorly characterized DNA repair pathway in prokaryotes (Wright *et al.*, 2018).

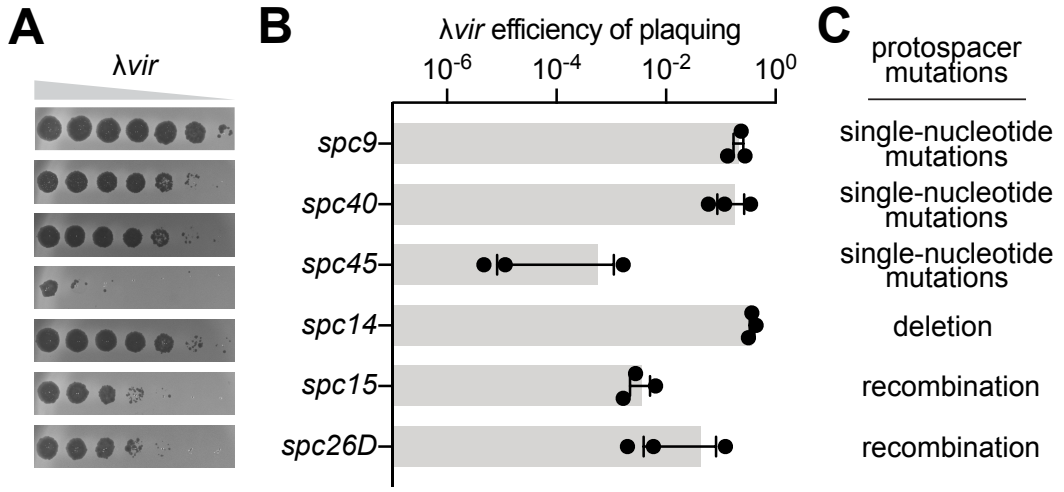


Figure 2.2 Phage λ*vir* escapes type II-A CRISPR-Cas targeting through the introduction of different target mutations

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir* on top agar plates seeded with *E. coli* expressing Cas9 programmed with different spacers. *spcNT*, non-targeting spacer. (**B**) Quantification of the data shown in (**A**) as the efficiency of plaquing relative to the non-targeting spacer. Mean ± SEM values of three independent experiments are shown. (**C**) Predominant target mutations detected after sequencing plaques obtained in (**A**).

Finally, escapers of *spc15* and *spc26D* contained multiple mutations across the target region that matched the sequence of cryptic prophages present in the chromosome of *E. coli* MG1655 (**Fig. 2.3**), suggesting homologous recombination with λ*vir* to generate escapers. Next, we sequenced the *spc9* target from phages present in plaques formed on lawns of non-targeting bacteria expressing Cas9 programmed with the non-targeting type II-A spacer *spcNT*. Given that Cas9 programmed with *spc9* leads to a reduction in plaque formation of less than one

order of magnitude, we should expect to detect about one inactivating mutation per ten targets sequenced. This was not the case, as no *spc9* target mutations were observed in 32 different plaques (data not shown). Therefore, the high numbers of phage mutants detected during *spc9*-mediated Cas9 targeting cannot be pre-existing and are likely generated during infection.

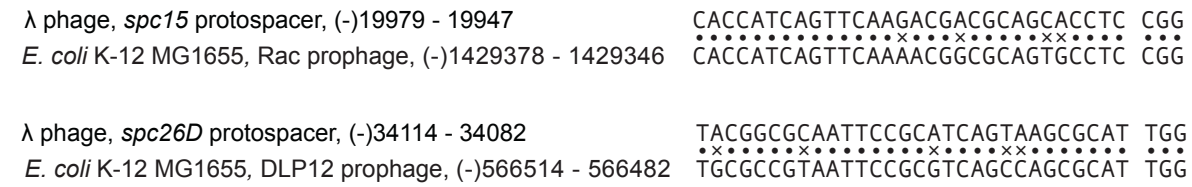


Figure 2.3 Phage escape through recombination with host chromosome

Alignment of *spc15* and *spc26D* targets in λ phage with homeologous cryptic prophage sequences in the *E. coli* MG1655 chromosome. Crosses indicate mismatches.

2.4 The phage λ Exo-Beta recombination system is required to generate a high frequency of Cas9 escape mutations

Given the mutation frequency for phage λ, calculated to be 7.7×10^{-8} mutations per base pair (Drake, 1991), the high frequency of escape mutations we observed for *spc9* and *spc40* ($<10^{-1}$) led us to hypothesize that mutations could be introduced during repair of the DSB generated by Cas9, by DNA polymerases that fill in ssDNA gaps across the cleaved sequence (Wright et al., 2018). The main pathway for DNA repair in *E. coli* is RecABCD recombination (Dillingham and Kowalczykowski, 2008). However, expression of Gam during λ infection will inhibit the RecBCD nuclease from

processing DSBs generated by Cas9 in the phage DNA. This allows the phage-encoded Exo-Beta recombination pathway to process DSBs during λ infection. To test if repair via the Exo-Beta pathway could mediate phage escape, we compared plaque formation between λ*vir* and λ*vir* Δ*exo* or λ*vir* Δ*bet* (**Fig. 2.4**) and found that deletion of the phage recombination genes severely decreased the number of escaper plaques.

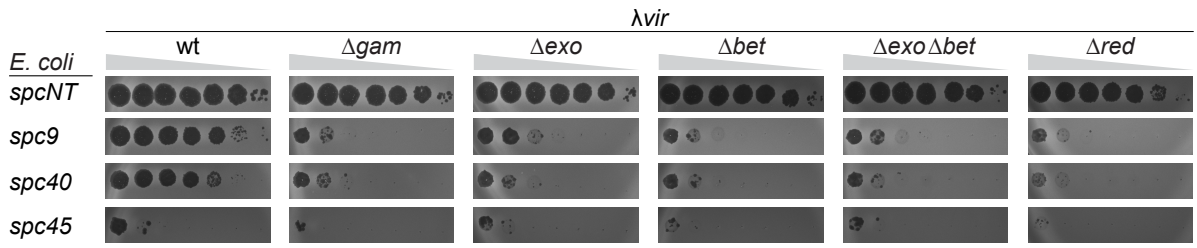


Figure 2.4 Deletions of λ *red* genes decrease the number of viral escapers

Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir* wild-type and mutant phages containing deletions within the *red* genes on top agar plates seeded with *E. coli* expressing Cas9 programmed with different spacers. *spcNT*, non-targeting spacer.

However, the reduction in escapers in the λ*vir* Δ*exo* and λ*vir* Δ*bet* phages could be a result of the involvement of this recombination system in the generation of target mutations, but also could be a consequence of the importance of these genes for λ replication (Echolas and Gingery, 1968; Signer and Weil, 1968). To determine the relative replication rates of these mutants, we performed quantitative PCR (qPCR) to measure the levels of phage DNA after 30 minutes of infection (**Fig. 2.5-AB**). Phages carrying the Δ*exo* Δ*bet* double deletion, as well as the elimination of the full Red system (Δ*red*), displayed a significant reduction in DNA content (**Fig. 2.5-B**), which

could limit the amount of λ DNA inside the host and thus lower the probability of mutagenic repair.

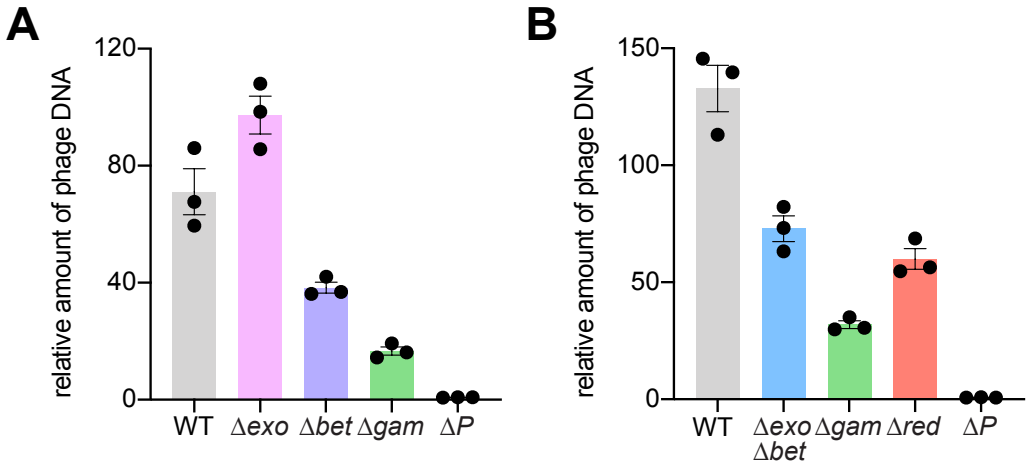


Figure 2.5 Effects of λ *red* gene deletions on phage DNA replication

(**A**) Quantitative PCR analysis of viral DNA within *E. coli* cells, 30 minutes after infection with different λ*vir* phages at an MOI of 1. Δ*P* values were obtained after infection with a non-replicative λ phage lacking the *P* gene necessary for initiating phage DNA replication. Fold-change values relative to λΔ*P* 15-min time point values are reported. Mean ± SEM values of three independent experiments are shown. (**B**) Same as (**A**) but analyzing different λ*vir* mutant phages.

Previous work has shown that a burst size defect of λ Δ*red* phages, presumably due to impaired formation of genome concatemers, can be overcome by a mutation that introduces a functional *E. coli chi* site (5'-GCTGGTGG-3') in the λ genome (Henderson and Weil, 1975). We wondered whether this modification would normalize the replication levels of the different λ*vir* mutants and therefore we introduced a *chi* site to generate λ*vir chi*^{1} (**Fig. 2.1-A**). First, we looked at the targeting efficiencies of the six spacers used in this study, finding almost identical results for λ*vir chi*^{1} (**Fig. 2.6**) to those obtained for λ*vir* (**Fig. 2.2**). Next, we generated

a Δ*exo* Δ*bet* double deletion variant of λ*vir chi*[1] and found that it accumulates similar DNA content during infection to the wild-type phage λ*vir chi*[1] (**Fig. 2.7**).

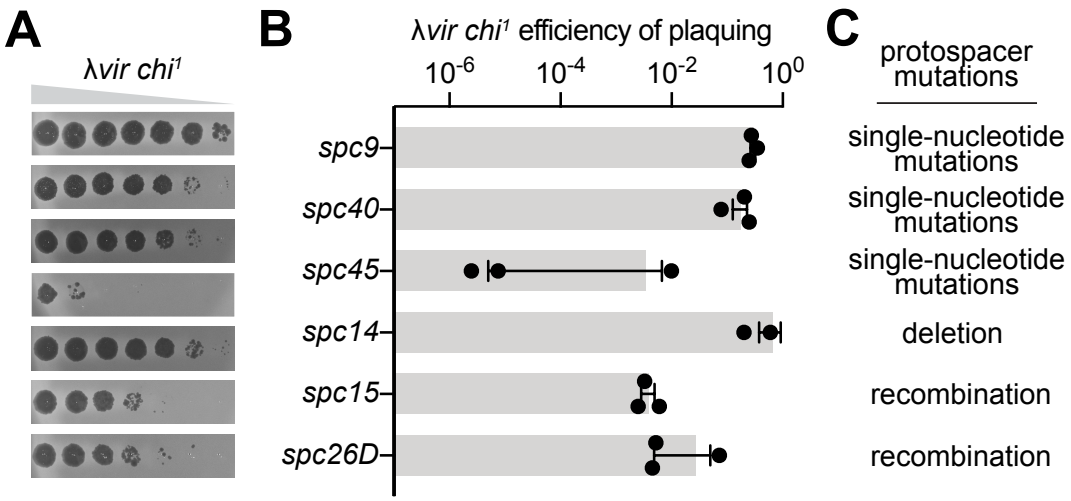


Figure 2.6 Type II-A CRISPR-Cas targeting of phage λ*vir chi*[1] is identical to targeting of phage λ*vir*

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*[1] on top agar plates seeded with *E. coli* expressing Cas9 programmed with different spacers. *spcNT*, non-targeting spacer. (**B**) Quantification of the data shown in (**A**) as the efficiency of plaquing relative to the non-targeting spacer. Mean ± SEM values of three independent experiments are shown. (**C**) Predominant target mutations detected after sequencing plaques obtained in (**A**).

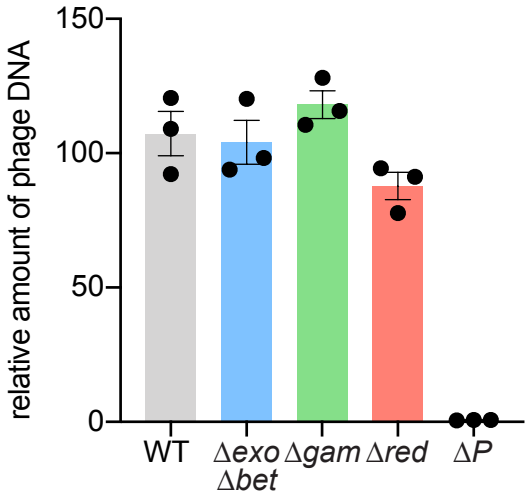


Figure 2.7 Effects of λ *red* gene deletions on phage λ*vir chi*[1] DNA replication

Quantitative PCR analysis of viral DNA within *E. coli* cells, 30 minutes after infection with different λ*vir chi*[1] phages at an MOI of 1. Δ*P* values were obtained after infection with a non-replicative λ phage lacking the *P* gene necessary for initiating phage DNA replication. Fold-change values relative to λΔ*P* 15-min time point values are reported. Mean ± SEM values of three independent experiments are shown.

We next used the λ*vir chi*[1] phage to test the contribution of Exo and Beta to the generation of escapers of *spc9*-, *spc40*- and *spc45*-mediated Cas9 targeting. We found very similar results to those obtained with the λ*vir* phage (compare **Fig 2.4** earlier and **Fig. 2.8** below). We found that the individual *exo* and *bet* gene deletions displayed the same phenotype as the Δ*exo* Δ*bet* double mutant (**Fig. 2.8**). The efficiency of plaquing was also quantified to highlight the large log-fold decreases in phage plaquing that occur during *spc9* and *spc40* targeting when Exo and Beta are deleted in the phage (**Fig. 2.9**).

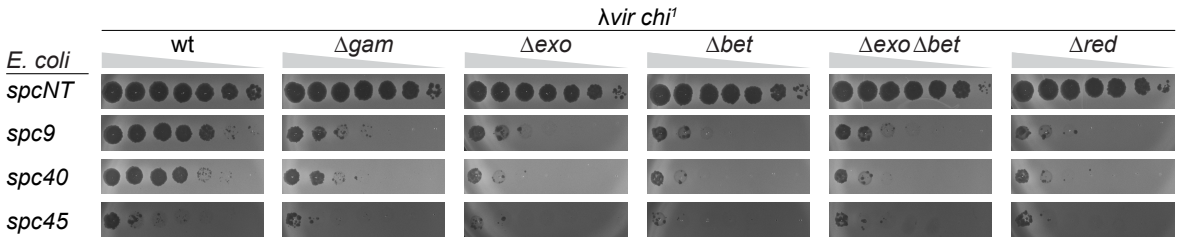


Figure 2.8 Deletions of λ *red* genes decrease the number of viral escapers during Cas9 targeting of phage λ*vir chi*[1]

Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*[1] wild-type and mutant phages containing deletions within the *red* genes on top agar plates seeded with *E. coli* expressing Cas9 programmed with different spacers. *spcNT*, non-targeting spacer.

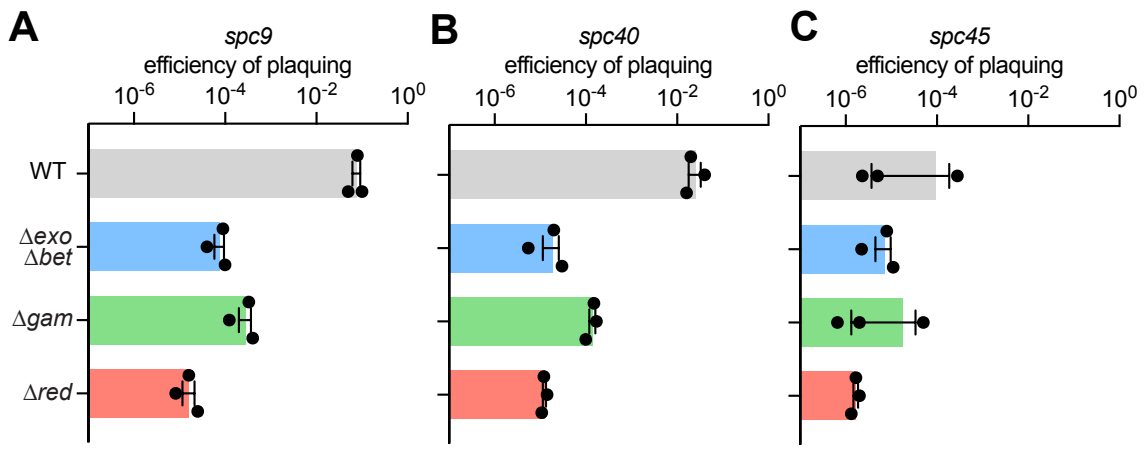


Figure 2.9 The phage λ Red system is required for the generation of Cas9 escape phages during type II-A CRISPR-Cas targeting

(A-C) Efficiency of plaquing of different λ*vir chi*[1] phages on lawns of *E. coli* expressing Cas9 programmed with *spc9* (**A**), *spc40* (**B**), or *spc45* (**C**). Mean ± SEM values of three independent experiments are shown.

Finally, we complemented the *spc9*-targeting *E. coli* host with plasmids expressing Exo or Beta and repeated the infections. The decrease in the number of escapers of the λ*vir chi*[1] Δ*exo* and Δ*bet* phages was rescued to λ*vir chi*[1] levels by the complementing plasmids (**Fig. 2.10** and **Fig. 2.11**). Since all the strains are infected with aliquots of the same viral population, this rescue demonstrates not only that *exo* and *bet* are responsible for the increase in the number of λ escapers, but also that escapers are not pre-existing but rather generated during infection. Altogether these results show that the Exo-Beta recombination system from phage λ promotes escape from Cas9 cleavage by increasing the production of phages carrying target site mutations.

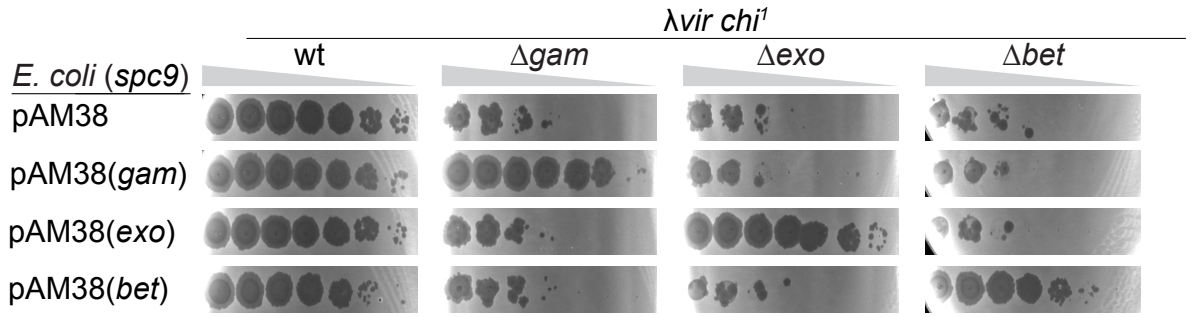


Figure 2.10 Genetic rescue of λ *red* genes in *red* mutant phages demonstrates that Red promotes evasion of CRISPR-Cas9 targeting

Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*[1] wild-type and mutant phages containing deletions within the *red* genes on top agar plates seeded with *E. coli* expressing Cas9 programmed with *spc9* and carrying the pAM38 vector expressing Exo, Beta or Gam.

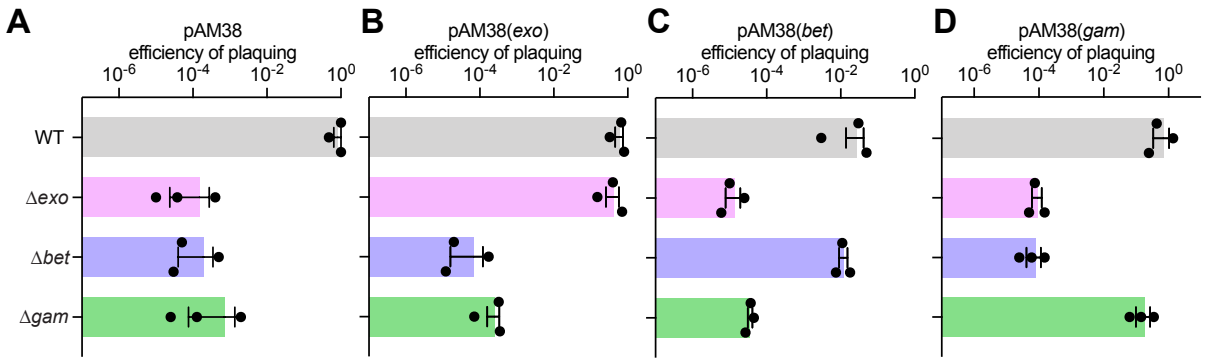


Figure 2.11 All three genes in the phage λ Red operon are necessary for CRISPR-Cas evasion and the generation of Cas9 escape phages

(A-D) Efficiency of plaquing of different λ*vir chi*[1] phages on lawns of *E. coli* expressing Cas9 programmed with *spc9* and carrying the pAM38 vector, either empty (**A**) or expressing Exo (**B**), Beta (**C**) or Gam (**D**) under arabinose induction. Mean ± SEM values of three independent experiments are shown for all measurements.

2.5 The λ Gam protein prevents RecBCD degradation of viral DNA

Next, we investigated the function of the third gene of the *red* operon, *gam*, during Cas9 targeting. We compared plaque formation by λ*vir* and λ*vir* Δ*gam* on lawns of cells expressing Cas9 and the *spc9*, *spc40* or *spc45* crRNA guides (**Fig. 2.4**). Compared to infection with λ*vir*, targeting of the Δ*gam* deletion mutant with *spc9* or *spc40* (but not with *spc45*) reduced plaque formation approximately three orders of magnitude. This reduction depended on the exonuclease activity of the RecBCD complex, as it was lost in *E. coli* Δ*recB* cells expressing $RecB^{D1080A}$, a nuclease-deficient version of RecB (Anderson et al., 1999), during *spc9* targeting of λ*vir* Δ*gam* (**Fig. 2.12**). Similar to the Δ*bet* and the Δ*exo* Δ*bet* mutants, absence of Gam resulted in a lower DNA content for λ*vir* (**Fig. 2.5**). We therefore repeated the experiment with the viruses containing one *chi* site, λ*vir chi*[1] and λ*vir chi*[1] Δ*gam*, whose DNA accumulation during infection are equivalent (**Fig. 2.7**). Quantification of plaque formation showed that the absence of Gam in λ*vir chi*[1] resulted in a reduction of two orders of magnitude for both *spc9* and *spc40* targeting (**Fig. 2.5-AB** and **Fig. 2.8**), which was rescued when the inhibitor was expressed from a plasmid in the *E. coli* host (**Fig. 2.10** and **Fig. 2.11**).

To investigate the effect of RecBCD activity on the phage genome targeted by Cas9, we performed next-generation sequencing (NGS) of cells harboring either *spc9* or *spcNT* infected for 25 minutes with either λ*vir chi*[1] or λ*vir chi*[1] Δ*gam*. When Gam was expressed, inactivation of RecBCD led to a dip in the reads spanning the *spc9*

target site, indicating the occurrence of only minimal degradation of phage DNA, most likely due to Exo-Beta resection and repair of the DSB and/or the action of other cellular nucleases at this location (**Fig. 2.13-A**). In contrast, the presence of active RecBCD in the host reduced the coverage to ~ 50% of the reads detected in the absence of Cas9 cleavage, across the whole λ genome, most likely due to destruction of the phage DNA by the RecBCD nuclease (**Fig. 2.13-B**). Interestingly, RecBCD activity did not seem to be impaired by the presence of the added *chi* site, and therefore we suspect that DNA degradation affected the overall replication of the phage as well. Altogether, these data show that the Gam protein protects the dsDNA ends generated by Cas9 cleavage from degradation by RecBCD, enabling Exo-Beta recombination to repair the DSB introduced during type II CRISPR-Cas targeting.

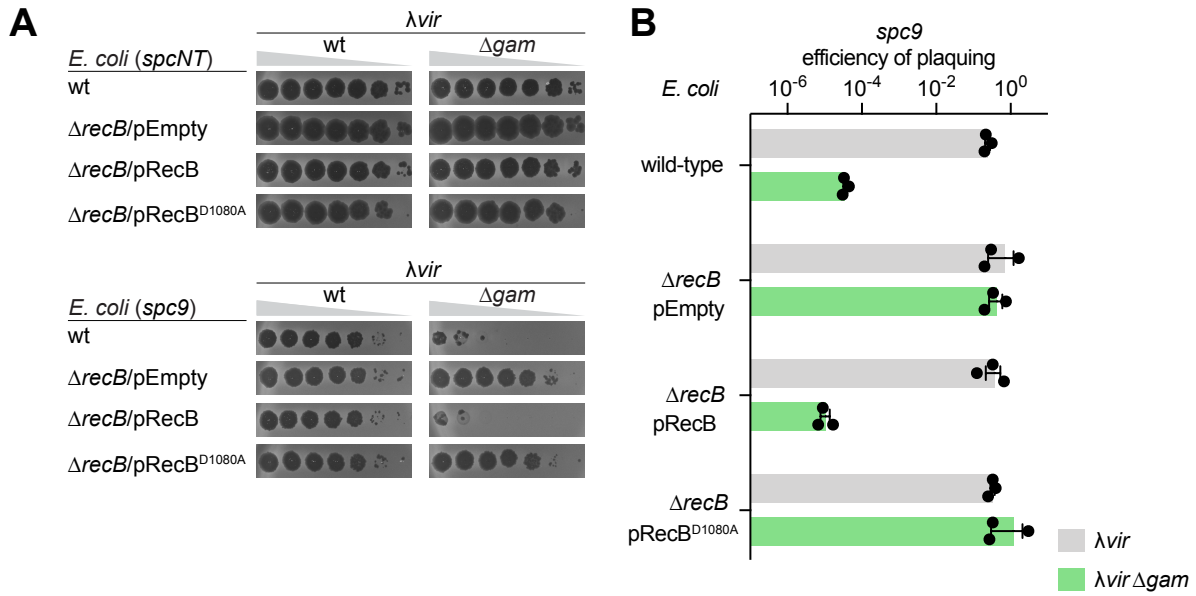


Figure 2.12 Gam protects phage λ against RecBCD following Cas9 cleavage

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir* or λ*vir* Δ*gam* phages on top agar plates seeded with *E. coli* expressing Cas9 programmed with *spcNT* or *spc9*, in the presence or absence of RecB. *spcNT*, non-targeting spacer. (**B**) Quantification of the data shown in (**A**) as the efficiency of plaquing relative to *spcNT*. Mean ± SEM values of three independent experiments are shown.

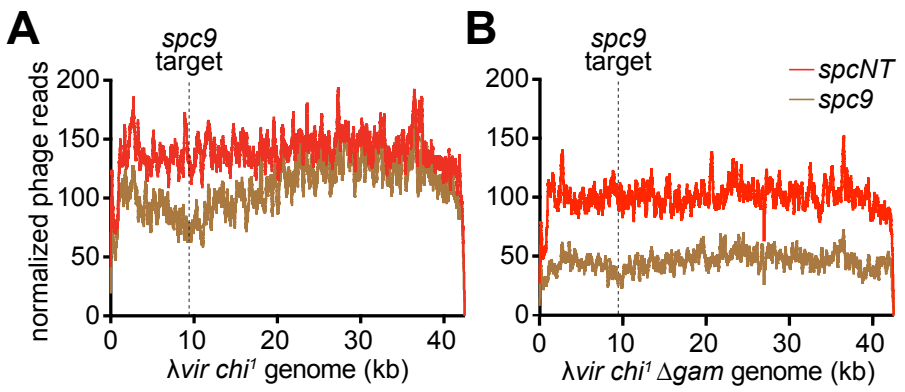


Figure 2.13 Gam prevents RecBCD degradation following Cas9 cleavage

(**A-B**) Normalized NGS reads of λ*vir* DNA, obtained 25 minutes after infection of *E. coli* expressing Cas9 programmed with *spc9* with either λ*vir chi*^{1} (**A**) or λ*vir chi*^{1} Δ*gam* (**B**) phages at an MOI of 5.

2.6 Repair of the λ phage genome by host RecABCD recombination generates a limited number of escapers

Lambdoid phages that do not carry RecBCD inhibitors contain multiple *chi* sites to prevent RecBCD degradation and promote RecA-mediated recombination (Bobay et al., 2013). To investigate how this pathway compares to Exo-Beta in the generation of Cas9 escape phages, we performed infections with λ*vir chi*^{1} Δ*red*, which harbors a deletion of all three components of the Red system as well as a *chi* site that stimulates RecABCD recombination. First, we confirmed that the Δ*red* deletion in λ*vir chi*^{1} does not significantly impact the replication of the phage (**Fig. 2.7**). We then measured escape frequencies for *spc9*-, *spc40*- and *spc45*-mediated Cas9 targeting and found that in the absence of the Red system the number of escapers was reduced by more than three orders of magnitude for *spc9* and *spc40* targeting but not significantly for *spc45* (**Fig. 2.8** and **Fig. 2.9**). These results suggest

that recombination through the RecABCD pathway is less efficient than Exo-Beta to generate escape mutations. We hypothesized that the chances of recombination and mutagenesis could be limited by the presence of only a single *chi* sequence, and therefore we introduced six additional *chi* sites into the λ*vir* Δ*red* genome (λ*vir chi*$^{2\text{-}7}$ Δ*red,* **Fig. 2.1-A**). The addition of the extra *chi* motifs did not severely impact viral DNA accumulation during infection with λ*vir* Δ*red* phages (**Fig. 2.14**). Critically, the number of Cas9 escaper plaques formed by λ*vir chi*$^{2\text{-}7}$ Δ*red* was not significantly different than the number generated by the λ*vir chi*^{1} Δ*red* phage (**Fig. 2.15**).

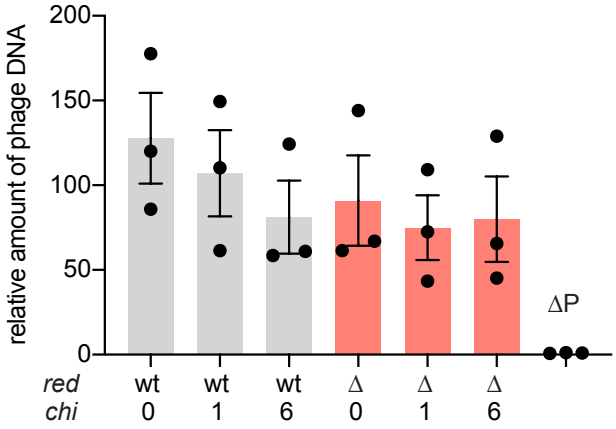


Figure 2.14 Effects of *chi* sites on λ*vir* DNA replication

Quantitative PCR analysis of viral DNA within *E. coli* cells, 30 minutes after infection with λ*vir* phages (differing in the presence or absence of the *red* operon and the number of *chi* sites) at an MOI of 1. Δ*P* values were obtained after infection with a non-replicative I phage lacking the *P* gene necessary for initiating phage DNA replication. Fold-change values relative to Δ*P* 30-min time point values are reported. Mean ± SEM values of three independent experiments are shown.

We also investigated escaper formation during infection of *E. coli* Δ*recD* hosts. In the absence of RecD, the RecBC complex lacks nuclease activity and instead unwinds dsDNA ends, constitutively loading RecA and promoting recombination

independently of *chi* recognition (a scenario somewhat equivalent to the presence of multiple *chi* sites flanking the DSB) (Anderson et al., 1997; Churchill et al., 1999). Using Δ*recD* cells, the number of *spc9* escaper plaques formed after infection with λ*vir chi*^{1} Δ*red* was more than two orders of magnitude higher than in the presence of RecD, but still one order of magnitude lower than in the presence of the Red system (**Fig. 2.16-AB**). Similar results were obtained using λ*vir chi*$^{2\text{-}7}$ phage (**Fig. 2.16-C**). Therefore, although RecABCD repair is not possible in a wild-type λ infection (Gam expression and the absence of *chi* sites prevent this), even in engineered conditions that maximize it, the number of escapers are significantly lower than those obtained through the repair of Cas9-generated DSBs by Exo-Beta.

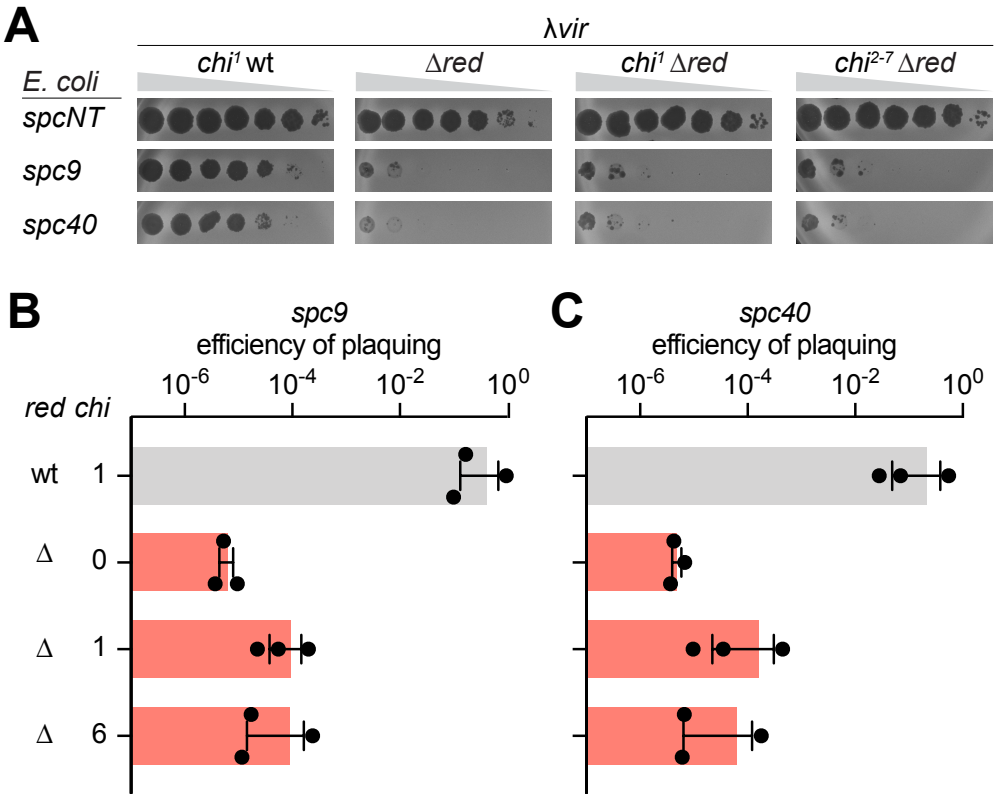


Figure 2.15 Effects of *chi* sites on λ Red protection against Cas9 targeting

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir* phages, differing in the presence or absence of the *red* operon and the number of *chi* sites, on top agar plates seeded with *E. coli* expressing Cas9 programmed with different spacers. *spcNT*, non-targeting spacer. (**B-C**) Quantification of the data shown in (**A**) as the efficiency of plaquing relative to *spcNT for spc9* (**B**) and *spc40* (**C**). Mean ± SEM values of three independent experiments are shown.

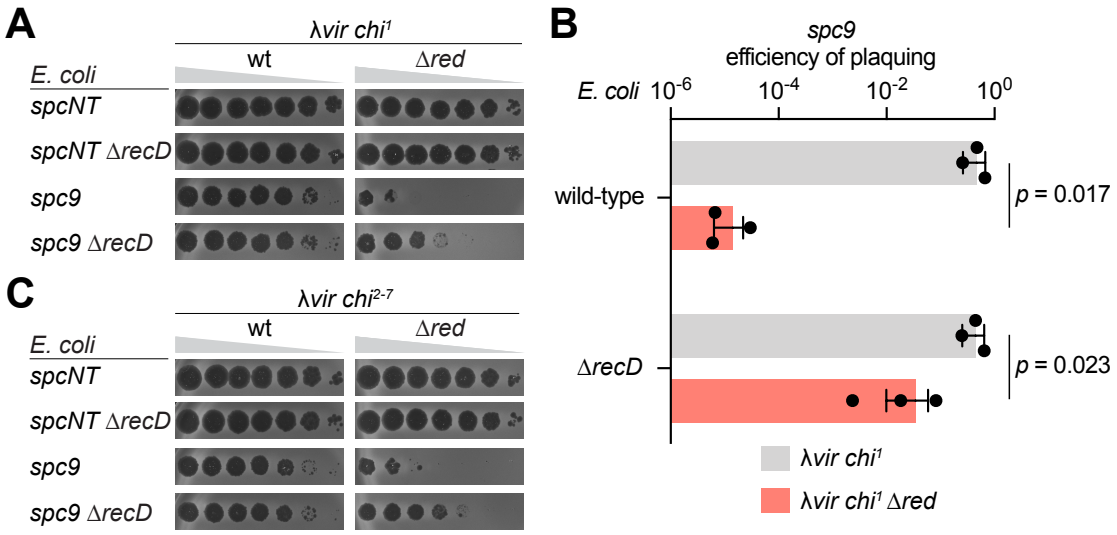


Figure 2.16 Constitutive RecABC recombination is less efficient than phage λ Red recombination in generating Cas9 escapers

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir* phages containing one *chi* site, on top agar plates seeded with *E. coli* expressing Cas9 programmed with different spacers, in the presence or absence of *recD*. *spcNT*, non-targeting spacer. (**B**) Quantification of the data shown in (**A**) as the efficiency of plaquing relative to *spcNT*. Mean ± SEM values of three independent experiments are shown. (**C**) Same as (**A**) but for phages containing *six* chi sites.

2.7 The Red recombination system is selected when λ phage is challenged with type II CRISPR-Cas targeting

Our results thus far suggest that the phage-encoded repair pathway is more effective than the host bacterial pathway at generating target mutations in the phage DNA, and more potently reduces the efficacy of type II CRISPR-Cas targeting. To

test this further, we performed a competition experiment between λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages on top agar (**Fig. 2.17**). Cas9-expressing *E. coli* cells carrying either *spcNT, spc9* or *spc45* targeting spacers were infected with a 1:1 mix of the two phages at a high multiplicity of infection (MOI) of 20. We found that, after checking for the presence of the *red* operon via PCR, all of the 36 plaques recovered after targeting by either of the tested spacers contained λ*vir chi*[1] phages, demonstrating a strong selection for viruses carrying the Red system. Altogether, our results reveal a two-pronged, post-cleavage strategy carried out by the Red system to overcome type II-A CRISPR-Cas targeting against phage λ, which confers a strong selective advantage to the bacteriophages carrying it: (i) Gam repression of RecBCD phage DNA degradation and recombination, and (ii) introduction of target mutations via Exo-Beta repair of Cas9-cleaved viral DNA.

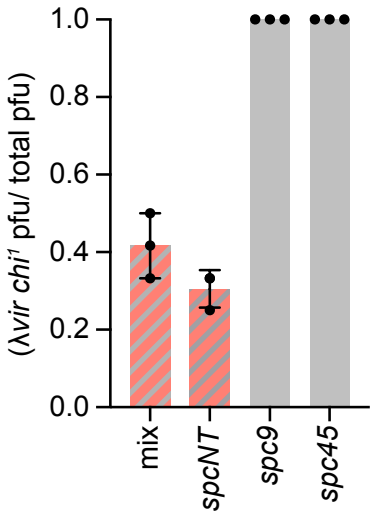


Figure 2.17 The phage λ Red system is selected during Cas9 targeting

Fraction of λ*vir chi*[1] from a total of 12 plaques obtained after co-infection with a 1:1 mixture of λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages of *E. coli* hosts harboring pCas9 programmed with a non-targeting spacer (*spcNT*), *spc9* or *spc45*, in top agar at a MOI of 20. Mean ± SD values of three independent experiments are shown.

CHAPTER 3. BACTERIOPHAGE RECOMBINATION SYSTEMS PROMOTE THE GENERATION OF CRISPR-CAS ESCAPE MUTATIONS

3.1 Background

CRISPR-Cas systems cleave DNA in a sequence-specific manner, generating free DNA ends that can be repaired, in theory, by DNA repair systems, including recombinational repair systems encoded by both bacteria and phages. In Chapter 2, we reported that the Red recombination system of coliphage λ mediates evasion of type II CRISPR-Cas targeting through repair of CRISPR-cleaved viral DNA. We find that this repair mediates immune evasion and promotes a higher frequency of Cas9 escape phages. In this chapter, we explore the different types of CRISPR target mutations facilitated by λ Red recombination and investigate how recombination may promote the rise of escape phages (Hossain et al., 2021). Finally, we test the generalizability of Red-mediated CRISPR-Cas evasion and escaper generation using the type I-E CRISPR-Cas immune system found in *E. coli* (Hossain et al., 2021).

3.2 The spectrum of Cas9 escape mutations differs in the presence or absence of λ Red recombination

Using the experimental setup from Chapter 2, we first sought to investigate whether the nature of mutations differed following RecBCD or Exo-Beta repair of Cas9 DSBs. To test this, we amplified and sequenced the targets of 12 λ*vir chi*^{1} or λ*vir chi*^{1}Δ*red* phage plaques that were able to evade *spc9*-mediated immunity (**Table**

3.1). We found that in the presence of Exo-Beta repair, 7/12 λ*vir chi*^{1} escapers displayed an adenine to cytosine change in position -5 (five nucleotides before the first nucleotide of the PAM) and the other 5/12 contained guanine to adenine mutations in the PAM. In contrast, in the absence of Exo-Beta repair, 10/12 λ*vir chi*^{1}Δ*red* escapers contained the -5A>C mutation, with only 1/12 harboring a change in the second guanine of the PAM to a cytosine (as opposed to adenine in the case of λ*vir chi*^{1} escapers). Similarly contrasting results were obtained for *spc40* escape phages (**Table 3.2**). In the case of *spc45* escape mutations, a trend showing seed mutations in 4/12 wild-type phage escapers (not present in the λ*vir chi*^{1}Δ*red* escapers in the plaques obtained in the experiment shown in **Fig. 2.8**, **Table 3.3**) was extended to 15/36 in the competition experiment of **Fig. 2.17** (**Table 3.4**).

To investigate the mutation spectrum of the *spc9* escapers in more detail, we subjected the target PCR products of hundreds of pooled escape plaques to NGS, in triplicate. Corroborating our analysis of individual plaques, most of the escape mutations comprised the seed sequence -5A>C substitution for both phages, and changes in the PAM region were markedly different in the absence of Red: 2G>A and 3G>A mutations were much less frequent and the 3G>T mutation was more abundant (**Fig. 3.1-A** and **Table 3.5**). Complementation with a plasmid expressing the *red* system showed that both the decrease in plaque formation (**Fig. 3.2**) as well as the mutation pattern of the λ*vir chi*^{1}Δ*red* phages were reverted to wild-type values (**Fig. 3.1-BC**, **Table 3.6,** and **Table 3.7**). These results illustrate the efficient rescue

of the Δ*red* deletion and demonstrate that not only the number of mutations, but their specific nature, are dictated by Exo-Beta repair after Cas9 cleavage.

Table 3.1 Sequences of CRISPR escape phages generated during *spc9* targeting λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages

E. coli host	phage	target	plaque	protospacer	PAM
wild-type	λ*vir chi* [1]	*spc9*	wild-type	AGATACCAGCTTCACGCTGGCGTGGATGCC	CGG
wild-type	λ*vir chi* [1]	*spc9*	escaper 1	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1]	*spc9*	escaper 2	AGATACCAGCTTCACGCTGGCGTGGATGCC	CGA
wild-type	λ*vir chi* [1]	*spc9*	escaper 3	AGATACCAGCTTCACGCTGGCGTGGATGCC	CGA
wild-type	λ*vir chi* [1]	*spc9*	escaper 4	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1]	*spc9*	escaper 5	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1]	*spc9*	escaper 6	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1]	*spc9*	escaper 7	AGATACCAGCTTCACGCTGGCGTGGATGCC	CGA
wild-type	λ*vir chi* [1]	*spc9*	escaper 8	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1]	*spc9*	escaper 9	AGATACCAGCTTCACGCTGGCGTGGATGCC	CGA
wild-type	λ*vir chi* [1]	*spc9*	escaper 10	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1]	*spc9*	escaper 11	AGATACCAGCTTCACGCTGGCGTGGATGCC	CGA
wild-type	λ*vir chi* [1]	*spc9*	escaper 12	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	wild-type	AGATACCAGCTTCACGCTGGCGTGGATGCC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	escaper 1	AGATACCAGCTTCACGCTGGCTTGGATTCC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	escaper 2	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	escaper 3	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	escaper 4	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	escaper 5	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	escaper 6	AGATACCAGCTTCACGCTGGCGTGGATGCC	CGC
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	escaper 7	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	escaper 8	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	escaper 9	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	escaper 10	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	escaper 11	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc9*	escaper 12	AGATACCAGCTTCACGCTGGCGTGGCTGCC	CGG

Table 3.2 Sequences of CRISPR escape phages generated during *spc40* targeting of λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages

E. coli host	phage	target	plaque	protospacer	PAM
wild-type	λ*vir chi* [1]	*spc40*	wild-type	AGGGATGCACCATTCTGAGATGTTTTTATT	TGG
wild-type	λ*vir chi* [1]	*spc40*	escaper 1	AGGGATGCACCATTCTGAGATGTTTTTGTT	TGG
wild-type	λ*vir chi* [1]	*spc40*	escaper 2	AGGGATGCACCATTCTGAGATGTTTTTGTT	TGG
wild-type	λ*vir chi* [1]	*spc40*	escaper 3	AGGGATGCACCATTCTGAGATGTTTTTATT	TAG
wild-type	λ*vir chi* [1]	*spc40*	escaper 4	AGGGATGCACCATTCTGAGATGTTTTTGTT	TGG
wild-type	λ*vir chi* [1]	*spc40*	escaper 5	AGGGATGCACCATTCTGAGATGTTTTTGTT	TGG
wild-type	λ*vir chi* [1]	*spc40*	escaper 6	AGGGATGCACCATTCTGAGATGTTTTTATT	TGC
wild-type	λ*vir chi* [1]	*spc40*	escaper 7	AGGGATGCACCATTCTGAGATGTTTTTGTT	TGG
wild-type	λ*vir chi* [1]	*spc40*	escaper 8	AGGGATGCACCATTCTGAGATGTTTTTATT	TAG
wild-type	λ*vir chi* [1] Δ*red*	*spc40*	wild-type	AGGGATGCACCATTCTGAGATGTTTTTATT	TGG
wild-type	λ*vir chi* [1] Δ*red*	*spc40*	escaper 1	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG
wild-type	λ*vir chi* [1] Δ*red*	*spc40*	escaper 2	AGGGATGCACCATTCTGAGATGTTTTTATT	TGT
wild-type	λ*vir chi* [1] Δ*red*	*spc40*	escaper 3	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG
wild-type	λ*vir chi* [1] Δ*red*	*spc40*	escaper 4	AGGGATGCACCATTCTGAGATGTTTTTGTT	TGG
wild-type	λ*vir chi* [1] Δ*red*	*spc40*	escaper 5	AGGGATGCACCATTCTGAGGTTTTTTTATT	TGG
wild-type	λ*vir chi* [1] Δ*red*	*spc40*	escaper 6	AGGGATGCACCATTCTGAGATGTTTTTGTT	TGG
wild-type	λ*vir chi* [1] Δ*red*	*spc40*	escaper 7	AGGGATGCACCATTCTGAGATTTTTTGTT	TGG
wild-type	λ*vir chi* [1] Δ*red*	*spc40*	escaper 8	AGGGATGCACCATTCTGAGATTTTTTATT	TGG

Table 3.3 Sequences of CRISPR escape phages generated during *spc45* targeting of λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages

E. coli host	phage	target	plaque	protospacer	PAM
wild-type	λ*vir chi* [1]	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
wild-type	λ*vir chi* [1]	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1]	*spc45*	escaper 2	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
wild-type	λ*vir chi* [1]	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
wild-type	λ*vir chi* [1]	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1]	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1]	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1]	*spc45*	escaper 7	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
wild-type	λ*vir chi* [1]	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1]	*spc45*	escaper 9	ACTCGTCAGGATGAATATTATCAAGCGGAA	AGG
wild-type	λ*vir chi* [1]	*spc45*	escaper 10	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1]	*spc45*	escaper 11	ACTCGTCAGGATGAATATTATCAAGCAGCA	AAG
wild-type	λ*vir chi* [1]	*spc45*	escaper 12	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 9	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 10	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 11	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 12	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG

Table 3.4 Sequences of CRISPR escape phages generated during *spc45* targeting in the competition experiment between λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages

Competition run#	*E. coli* host	phage	target	plaque	protospacer	PAM
1	wild-type	λ*vir chi* [1]	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
1	wild-type	λ*vir chi* [1]	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGC
1	wild-type	λ*vir chi* [1]	*spc45*	escaper 2	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
1	wild-type	λ*vir chi* [1]	*spc45*	escaper 3	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
1	wild-type	λ*vir chi* [1]	*spc45*	escaper 4	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
1	wild-type	λ*vir chi* [1]	*spc45*	escaper 5	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
1	wild-type	λ*vir chi* [1]	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
1	wild-type	λ*vir chi* [1]	*spc45*	escaper 7	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
1	wild-type	λ*vir chi* [1]	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
1	wild-type	λ*vir chi* [1]	*spc45*	escaper 9	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
1	wild-type	λ*vir chi* [1]	*spc45*	escaper 10	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
1	wild-type	λ*vir chi* [1]	*spc45*	escaper 11	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
1	wild-type	λ*vir chi* [1]	*spc45*	escaper 12	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
2	wild-type	λ*vir chi* [1]	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
2	wild-type	λ*vir chi* [1]	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
2	wild-type	λ*vir chi* [1]	*spc45*	escaper 2	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
2	wild-type	λ*vir chi* [1]	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	ACG
2	wild-type	λ*vir chi* [1]	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
2	wild-type	λ*vir chi* [1]	*spc45*	escaper 5	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
2	wild-type	λ*vir chi* [1]	*spc45*	escaper 6	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
2	wild-type	λ*vir chi* [1]	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	ACG
2	wild-type	λ*vir chi* [1]	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGA
2	wild-type	λ*vir chi* [1]	*spc45*	escaper 9	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
2	wild-type	λ*vir chi* [1]	*spc45*	escaper 10	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
2	wild-type	λ*vir chi* [1]	*spc45*	escaper 11	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
2	wild-type	λ*vir chi* [1]	*spc45*	escaper 12	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGC
3	wild-type	λ*vir chi* [1]	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
3	wild-type	λ*vir chi* [1]	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
3	wild-type	λ*vir chi* [1]	*spc45*	escaper 2	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
3	wild-type	λ*vir chi* [1]	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
3	wild-type	λ*vir chi* [1]	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
3	wild-type	λ*vir chi* [1]	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
3	wild-type	λ*vir chi* [1]	*spc45*	escaper 6	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
3	wild-type	λ*vir chi* [1]	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
3	wild-type	λ*vir chi* [1]	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	ACG
3	wild-type	λ*vir chi* [1]	*spc45*	escaper 9	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
3	wild-type	λ*vir chi* [1]	*spc45*	escaper 10	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
3	wild-type	λ*vir chi* [1]	*spc45*	escaper 11	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
3	wild-type	λ*vir chi* [1]	*spc45*	escaper 12	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT

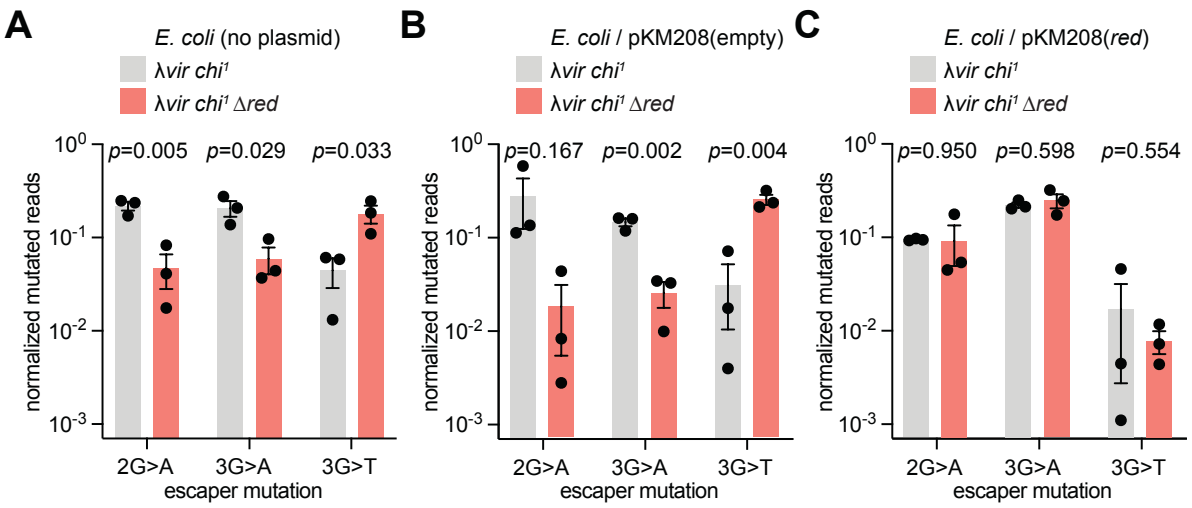


Figure 3.1 Phages that escape Cas9 cleavage exhibit specific target mutations in the presence of λ Red

(**A-C**) Normalized mutant NGS reads of the *spc9* targets of λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages that escape type II-A targeting on lawns of *E. coli* expressing Cas9 programmed with *spc9* and no additional plasmid (**A**), the pKM208(empty) vector (**B**) or the same vector expressing the λ Red system (**C**). Mean ± SEM values of three independent experiments are shown.

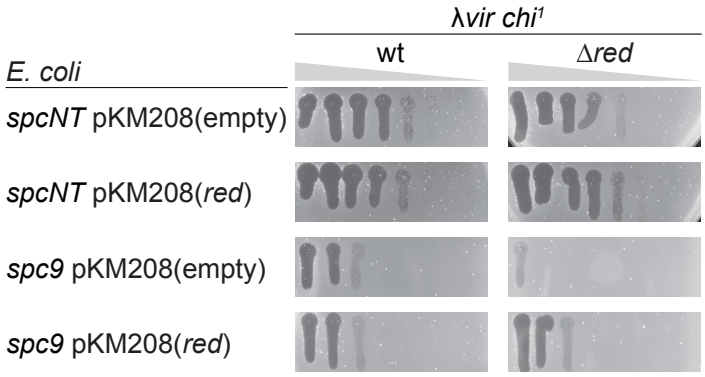


Figure 3.2 Expression of the λ *red* operon enables phage escape during Cas9 targeting

Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*[1] or λ*vir chi*[1] Δ*red* phages on top agar plates seeded with *E. coli* expressing Cas9 programmed with *spcNT* or *spc9* and carrying the pKM208 vector either empty or expressing the λ Red system under 1 mM IPTG induction. *spcNT*, non-targeting spacer.

Table 3.5 Normalized mutation reads obtained after next-generation sequencing of *spc9* escape phages

E. coli	Plasmid 1	Plasmid 2	Phage	Mutation	Replicate 1	Replicate 2	Replicate 3	Average	SEM	*p*-value
wild-type	pCas9::*spc9*	no plasmid	λ*vir chi* ^{1}	+2 G→A	0.24761083	0.23644511	0.17209373	0.218716557	0.0235332	
wild-type	pCas9::*spc9*	no plasmid	λ*vir chi* ^{1} Δ*red*	+2 G→A	0.01757311	0.08298657	0.04097491	0.047178197	0.01913627	0.004817
wild-type	pCas9::*spc9*	no plasmid	λ*vir chi* ^{1}	+3 G→A	0.20773858	0.27561453	0.13751141	0.20695484	0.03986886	
wild-type	pCas9::*spc9*	no plasmid	λ*vir chi* ^{1} Δ*red*	+3 G→A	0.04411765	0.03712575	0.09686496	0.059369453	0.01885609	0.028665
wild-type	pCas9::*spc9*	no plasmid	λ*vir chi* ^{1}	+3 G→T	0.06122841	0.05878315	0.01313247	0.044381343	0.01564037	
wild-type	pCas9::*spc9*	no plasmid	λ*vir chi* ^{1} Δ*red*	+3 G→T	0.24644621	0.18435248	0.10974843	0.180182373	0.0395163	0.033045

Table 3.6 Normalized mutation reads obtained after next-generation sequencing of *spc9* escape phages formed during infection of hosts carrying the pKM208(empty) plasmid

E. coli	Plasmid 1	Plasmid 2	Phage	Mutation	Replicate 1	Replicate 2	Replicate 3	Average	SEM	*p*-value
wild-type	pCas9::*spc9*	pKM208(empty)	λ*vir chi* ^{1}	+2 G→A	0.55993448	0.13039021	0.10845687	0.26626052	0.14697343	
wild-type	pCas9::*spc9*	pKM208(empty)	λ*vir chi* ^{1} Δ*red*	+2 G→A	0.00799062	0.04209563	0.00267865	0.0175883	0.01234924	0.167072
wild-type	pCas9::*spc9*	pKM208(empty)	λ*vir chi* ^{1}	+3 G→A	0.11332602	0.15570561	0.15333804	0.14078989	0.01374893	
wild-type	pCas9::*spc9*	pKM208(empty)	λ*vir chi* ^{1} Δ*red*	+3 G→A	0.03151316	0.03293335	0.0095639	0.024670137	0.00756424	0.001779
wild-type	pCas9::*spc9*	pKM208(empty)	λ*vir chi* ^{1}	+3 G→T	0.0038255	0.01689944	0.06893461	0.029886517	0.01988548	
wild-type	pCas9::*spc9*	pKM208(empty)	λ*vir chi* ^{1} Δ*red*	+3 G→T	0.20503703	0.22672721	0.30633577	0.246033337	0.0307945	0.004138

Table 3.7 Normalized mutation reads obtained after next-generation sequencing of *spc9* escape phages formed during infection of hosts carrying the pKM208(*red*) plasmid, which expresses the λ *red* operon

E. coli	Plasmid 1	Plasmid 2	Phage	Mutation	Replicate 1	Replicate 2	Replicate 3	Average	SEM	*p*-value
wild-type	pCas9::*spc9*	pKM208(*red*)	λ*vir chi* ^{1}	+2 G→A	0.08964586	0.09143814	0.09469538	0.09192646	0.00147798	
wild-type	pCas9::*spc9*	pKM208(*red*)	λ*vir chi* ^{1} Δ*red*	+2 G→A	0.17157427	0.04363428	0.05231526	0.089174603	0.04127598	0.950076
wild-type	pCas9::*spc9*	pKM208(*red*)	λ*vir chi* ^{1}	+3 G→A	0.20566482	0.24212564	0.19639183	0.21472743	0.01395819	
wild-type	pCas9::*spc9*	pKM208(*red*)	λ*vir chi* ^{1} Δ*red*	+3 G→A	0.31267748	0.16825234	0.23861263	0.239847483	0.04169652	0.598371
wild-type	pCas9::*spc9*	pKM208(*red*)	λ*vir chi* ^{1}	+3 G→T	0.04468712	0.00106689	0.00430816	0.01668739	0.0140311	
wild-type	pCas9::*spc9*	pKM208(*red*)	λ*vir chi* ^{1} Δ*red*	+3 G→T	0.00700691	0.01136689	0.0042523	0.007542033	0.00207116	0.554172

3.3 The error-prone DNA polymerase Pol IV is important to generate Cas9 escape mutations

To investigate the origin of the mutations, we evaluated the involvement of the error-prone polymerases Pol II, IV and V (Henrikus et al., 2018), since they participate in the synthesis of the DNA required to fill in the gaps present in the recombination products (Wright et al., 2018). We compared the number of plaques obtained after infection with λ*vir chi*^{1} of *E. coli* hosts harboring *spc9* but with different

mutant backgrounds: Δ*polB* (Pol II), Δ*dinB* (Pol IV) and Δ*umuC* (Pol V). While the elimination of Pol II and V activity did not impact the generation of *spc9* escapers, absence of Pol IV reduced the number of plaques formed by Exo-Beta repair of the λ*vir chi*^{1} phage by approximately two orders of magnitude (**Fig. 3.3-AB**). Although the quantification of λ*vir chi*^{1}Δ*red* phages that escape *spc9* targeting was close to our limit of detection, the number of plaques obtained after infection of Δ*dinB* hosts was much closer to the numbers obtained in wild-type, Δ*polB* and Δ*umuC* cells (**Fig. 3.3-CD**). The distribution of the 2G>A, 3G>A and 3G>T mutations, however, did not change in the presence or absence of Pol IV (**Fig. 3.4** and **Table 3.8**). In addition, we found that overexpression of *dinB* from a plasmid (pDinB) did not increase the number of *spc9* λ*vir chi*^{1}Δ*red* escapers (**Fig. 3.5**). Altogether, these results indicate that while Pol IV is the most important error-prone DNA polymerase for the increase of phage escapers during Exo-Beta repair of Cas9-cleaved λ DNA, the activity of this polymerase is not sufficient and needs to act together with the Red system to introduce specific escape mutations.

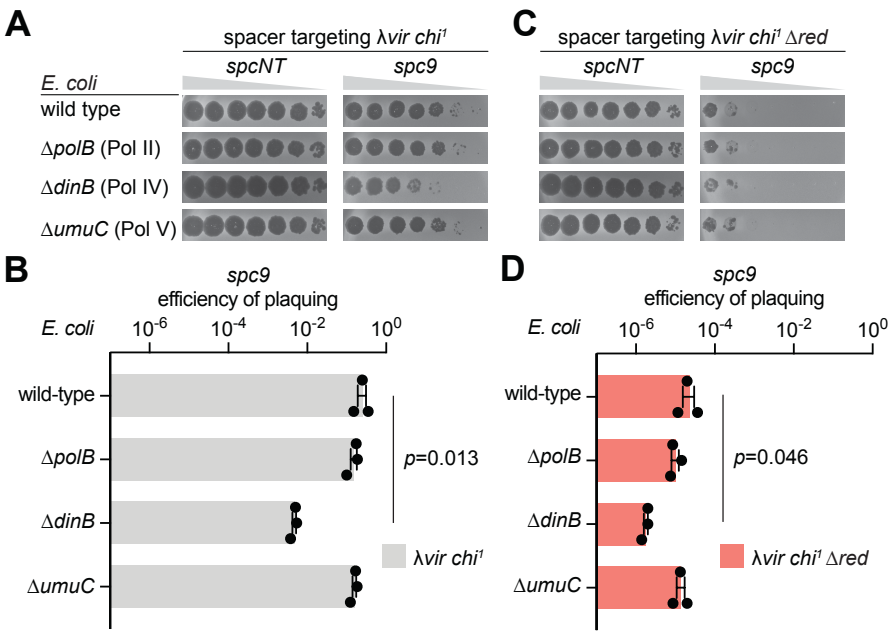


Figure 3.3 The error-prone DNA polymerase Pol IV is important for the generation of Cas9 escape phages during *spc9* targeting

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*^{1} on top agar plates seeded with *E. coli* expressing Cas9 programmed with a non-targeting spacer (*spcNT*) or *spc9* and carrying deletions in the genes encoding Pol II (*polB*), Pol IV (*dinB*) or Pol V (*umuC*). (**B**) Efficiency of plaquing of λ*vir chi*^{1} on lawns of *E. coli* expressing Cas9 programmed with *spc9*, in the presence or absence of genes encoding different *E. coli* error-prone DNA polymerases. Mean ± SEM values of three independent experiments are shown. (**C**) Same as (**A**) but with λ*vir chi*^{1} Δ*red*. (**D**) Same as (**B**) but with λ*vir chi*^{1} Δ*red*.

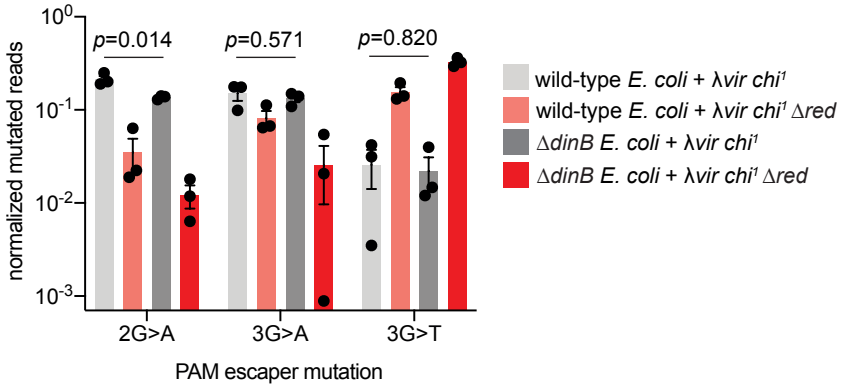


Figure 3.4 Pol IV does not affect the spectrum of Cas9 escape mutations

Normalized mutant NGS reads of the *spc9* targets of λ*vir chi*^{1} or λ*vir chi*^{1} Δ*red* phages that escape type II-A targeting on lawns of *E. coli* expressing Cas9 programmed with *spc9* in the presence or absence of Pol IV (*dinB*). Mean ± SEM values of three independent experiments are shown.

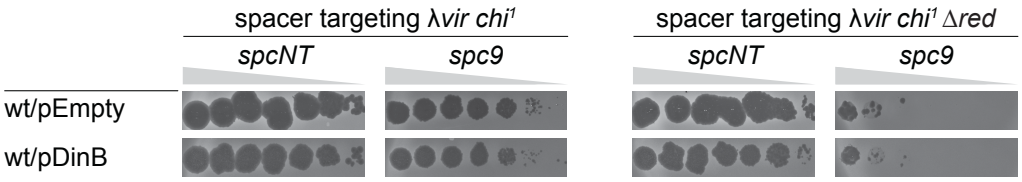


Figure 3.5 Pol IV overexpression does not increase phage escape from Cas9 targeting

Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*[1] or λ*vir chi*[1] Δ*red* phages on top agar plates seeded with *E. coli* expressing Cas9 programmed with *spcNT* or *spc9* and carrying a pDinB plasmid, expressing Pol IV, or a vector control (pEmpty). *spcNT*, non-targeting spacer.

Table 3.8 Normalized mutation reads obtained after next-generation sequencing of *spc9* escape phages in the presence or absence of Pol IV (*dinB*)

E. coli	Plasmid 1	Plasmid 2	Phage	Mutation	Replicate 1	Replicate 2	Replicate 3	Average	SEM	*p*-value
wild-type	pCas9::*spc9*	no plasmid	λ*vir chi*[1]	+2 G→A	0.24955775	0.20084975	0.19022799	0.213545163	0.0182655	
Δ*dinB*	pCas9::*spc9*	no plasmid	λ*vir chi*[1]	+2 G→A	0.13862261	0.12900132	0.1410767	0.136233543	0.00368485	0.014271
wild-type	pCas9::*spc9*	no plasmid	λ*vir chi*[1] Δ*red*	+2 G→A	0.06344118	0.01898764	0.0223332	0.034920673	0.01429292	
Δ*dinB*	pCas9::*spc9*	no plasmid	λ*vir chi*[1] Δ*red*	+2 G→A	0.00636114	0.01810365	0.01188467	0.012116487	0.00339175	0.195525
wild-type	pCas9::*spc9*	no plasmid	λ*vir chi*[1]	+3 G→A	0.17521819	0.17731563	0.0991972	0.150577007	0.02569704	
Δ*dinB*	pCas9::*spc9*	no plasmid	λ*vir chi*[1]	+3 G→A	0.14936349	0.10952025	0.14029025	0.133057997	0.01205681	0.570507
wild-type	pCas9::*spc9*	no plasmid	λ*vir chi*[1] Δ*red*	+3 G→A	0.06709006	0.06455563	0.1127768	0.081474163	0.01566841	
Δ*dinB*	pCas9::*spc9*	no plasmid	λ*vir chi*[1] Δ*red*	+3 G→A	0.05441882	0.02060062	0.00088989	0.02530311	0.01563033	0.064113
wild-type	pCas9::*spc9*	no plasmid	λ*vir chi*[1]	+3 G→T	0.04194669	0.00350624	0.03153471	0.025662547	0.01147866	
Δ*dinB*	pCas9::*spc9*	no plasmid	λ*vir chi*[1]	+3 G→T	0.0396359	0.01207327	0.01474779	0.02215232	0.00877582	0.820002
wild-type	pCas9::*spc9*	no plasmid	λ*vir chi*[1] Δ*red*	+3 G→T	0.19499482	0.13090558	0.14128162	0.15572734	0.01986091	
Δ*dinB*	pCas9::*spc9*	no plasmid	λ*vir chi*[1] Δ*red*	+3 G→T	0.2934355	0.36305414	0.32428948	0.326926373	0.02014037	0.003761

3.4 Inefficient Cas9 cleavage increases the generation of escape mutations

Our model for the generation of escape mutations requires cleavage of a target sequence before it can be repaired by Exo-Beta recombination. To test this, we performed experiments using dCas9, a Cas9 mutant that does not cleave its target DNA but can bind it and interrupt transcription (Bikard et al., 2013; Qi et al., 2013). When programmed with *spc40*, dCas9 reduced λ*vir chi*[1] plaque formation by more than one order of magnitude (presumably by preventing phage gene expression),

similar to the level of immunity it provided in hosts expressing wild-type Cas9. In contrast, we did observe a further decrease in plaques during infection with the λ*vir chi*^{1}Δ*red* phage (**Fig. 3.6**). Importantly, the target sequences of the dCas9 escapers, both in the presence and absence of the Red system during infection, resembled those of the Cas9 escapers in the absence of Red (**Table 3.2** and **Table 3.9**). Collectively, these results demonstrate that Cas9 cleavage is necessary for the generation of escape phages through the Exo-Beta recombination pathway.

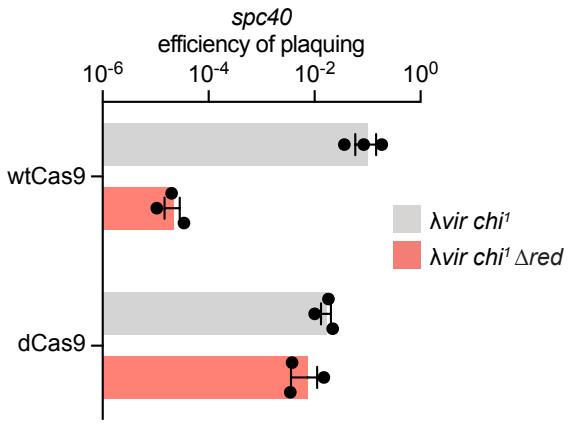


Figure 3.6 Cas9 cleavage is necessary for the generation of escape phages through the λ Red recombination pathway

Efficiency of plaquing of λ*vir chi*^{1} or λ*vir chi*^{1} Δ*red* phages on lawns of *E. coli* expressing Cas9 or dCas9 programmed with *spc40*. Mean ± SEM values of three independent experiments are shown.

Table 3.9 Sequences of CRISPR escape phages generated during *spc40* targeting of λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages using nuclease-dead dCas9

E. coli host	phage	Cas9 mutations	target	plaque	protospacer	PAM
wild-type	λ*vir chi* [1]	D10A, H840A	*spc40*	wild-type	AGGGATGCACCATTCTGAGATGTTTTTATT	TGG
wild-type	λ*vir chi* [1]	D10A, H840A	*spc40*	escaper 1	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG
wild-type	λ*vir chi* [1]	D10A, H840A	*spc40*	escaper 2	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG
wild-type	λ*vir chi* [1]	D10A, H840A	*spc40*	escaper 3	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG
wild-type	λ*vir chi* [1]	D10A, H840A	*spc40*	escaper 4	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG
wild-type	λ*vir chi* [1]	D10A, H840A	*spc40*	escaper 5	AGGGATGCACCATTCTGAGATGTTTTTGTT	TGG
wild-type	λ*vir chi* [1]	D10A, H840A	*spc40*	escaper 6	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG
wild-type	λ*vir chi* [1]	D10A, H840A	*spc40*	escaper 7	AGGGATGCACCATTCTGAGATGTTTTTGTT	TGG
wild-type	λ*vir chi* [1]	D10A, H840A	*spc40*	escaper 8	AGGGATGCACCATTCTGAGATGTTTTTATT	TAG
wild-type	λ*vir chi* [1] Δ*red*	D10A, H840A	*spc40*	wild-type	AGGGATGCACCATTCTGAGATGTTTTTATT	TGG
wild-type	λ*vir chi* [1] Δ*red*	D10A, H840A	*spc40*	escaper 1	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG
wild-type	λ*vir chi* [1] Δ*red*	D10A, H840A	*spc40*	escaper 2	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG
wild-type	λ*vir chi* [1] Δ*red*	D10A, H840A	*spc40*	escaper 3	AGGGATGCACCATTCTGAGATGTTTTTACT	TGG
wild-type	λ*vir chi* [1] Δ*red*	D10A, H840A	*spc40*	escaper 4	AGGGATGCACCATTCTGAGATGTTTTTATT	TGC
wild-type	λ*vir chi* [1] Δ*red*	D10A, H840A	*spc40*	escaper 5	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG
wild-type	λ*vir chi* [1] Δ*red*	D10A, H840A	*spc40*	escaper 6	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG
wild-type	λ*vir chi* [1] Δ*red*	D10A, H840A	*spc40*	escaper 7	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG
wild-type	λ*vir chi* [1] Δ*red*	D10A, H840A	*spc40*	escaper 8	AGGGATGCACCATTCTGAGATTTTTTTATT	TGG

Repair of a DSB requires recombination with an intact copy of the viral genome. Therefore, under conditions where cleavage is efficient and most uncut viral genomes are eliminated from the host cell, escaper generation through Red-based recombination should be inhibited. We wondered whether this was the case during *spc45*-mediated immunity, in which escaper generation was severely restricted for both λ*vir chi*[1] and λ*vir chi*[1]Δ*red* phages (**Fig. 2.8** and **Fig. 2.9-C**). We attempted to decrease the cleavage efficiency of Cas9 by mutating the *spc45* sequence, introducing a mismatch between the crRNA produced by this spacer (*spc45c*) and the seed sequence (Jiang et al., 2013; Jinek et al., 2012) of its target in the λ genome (**Fig. 3.7-A**). NGS of DNA extracted from cells 25 minutes after infection with λ*vir chi*[1] showed that indeed, *spc45c* targeting resulted in an intermediate number of reads across the genome, higher than those for *spc45* targeting, but lower than those for

non-targeting, conditions (**Fig. 3.7-B**). These differences in phage DNA accumulation correlated with the immunity provided by each spacer (**Fig. 3.7-C**). More important, when we infected cells harboring *spc45c* with λ*vir chi*^{1}Δ*red* phage, escapers were reduced by almost three orders of magnitude compared to wild-type λ*vir chi*^{1} (**Fig. 3.7-C**). These results corroborate a requirement for a high number of uncleaved phage genomes for the generation of Cas9 escapers through Exo-Beta repair.

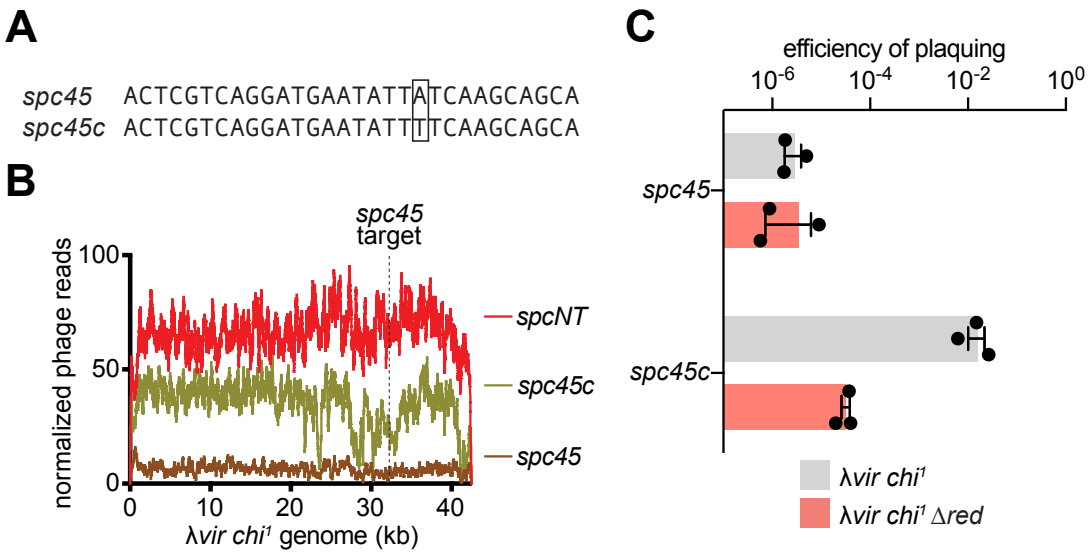


Figure 3.7 Inefficient Cas9 cleavage is required for the generation of high numbers of escape phages

(**A**) Mutation in the seed sequence of *spc45* to generate *spc45c*. (**B**) Normalized NGS reads of λ*vir* DNA, obtained 25 minutes after infection of *E. coli* expressing Cas9 programmed with *spcNT*, *spc45* or *spc45c* with λ*vir chi*^{1} phage at an MOI of 5. *spcNT*, non-targeting spacer. (**C**) Efficiency of plaquing of λ*vir chi*^{1} or λ*vir chi*^{1} Δ*red* phages on lawns of *E. coli* expressing Cas9 programmed with *spc45* or *spc45c*. Mean ± SEM values of three independent experiments are shown.

We also investigated the effects of the different Cas9 cleavage efficiencies in the dynamics of escaper generation during infection. To do this, we performed infections in liquid media since, as opposed to the genotyping of plaques on bacterial

lawns in which non-mutant phages are competed out over the 24-hour plate incubation period, these can reveal short-term variations in the appearance of target mutations. We infected liquid cultures expressing Cas9 programmed with the strong *spc45* or the weak *spc45c* crRNAs, with either λ*vir chi*^{1}or λ*vir chi*^{1}Δ*red* phage. We took culture aliquots at 50-minute intervals after infection (the average duration of a lytic cycle for λ*vir chi*^{1}, **Fig. 3.8**) and plated them onto a non-CRISPR strain to obtain phage plaques (**Fig. 3.9-A**). We then analyzed the target genotype of 8 plaques per time point (**Table 3.10-Table 3.13**). After infection with both λ*vir chi*^{1} and λ*vir chi*^{1}Δ*red* phages, *spc45* escapers were rapidly detected, with all of the eight sequenced plaques formed by mutant phage by 200 minutes (**Fig. 3.9-B**, **Table 3.10** and **Table 3.11**). In contrast, λ*vir chi*^{1} phages harboring wild-type *spc45c* targets continued to be detected over the nine hours of the experiment (**Fig. 3.9-B** and **Table 3.12**). Importantly, in the absence of λ Red recombination during the weak *spc45c*-mediated immunity, the rise of escapers was similar to that observed in *spc45* cultures (**Fig. 3.9-B** and **Table 3.13**). These results demonstrate that the generation of escaper phages through Exo-Beta recombination during poor immunity is a relatively slow process, probably preceded by multiple rounds of cleavage and repair that regenerate wild-type phages with intact targets that can continue propagating. On the other hand, our observations suggest that the rapid accumulation of escaper mutations in conditions where Exo-Beta repair is limited due to efficient target cleavage is most likely driven by the presence of pre-existing target mutations within the viral population.

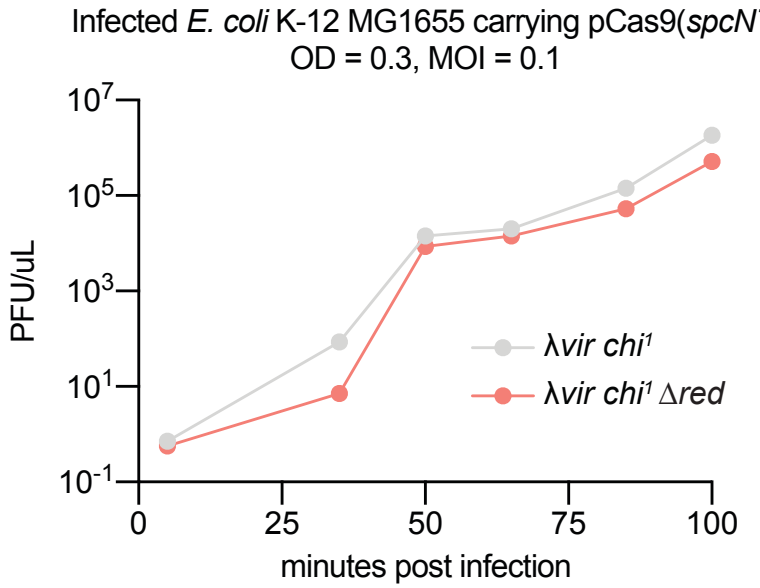


Figure 3.8 Wild-type and *red* mutant phages propagate similarly over time

Viral infection assay to measure the propagation of λ*vir chi*^{1} or λ*vir chi*^{1} Δ*red* phages in *E. coli* over time. PFU, plaque forming units.

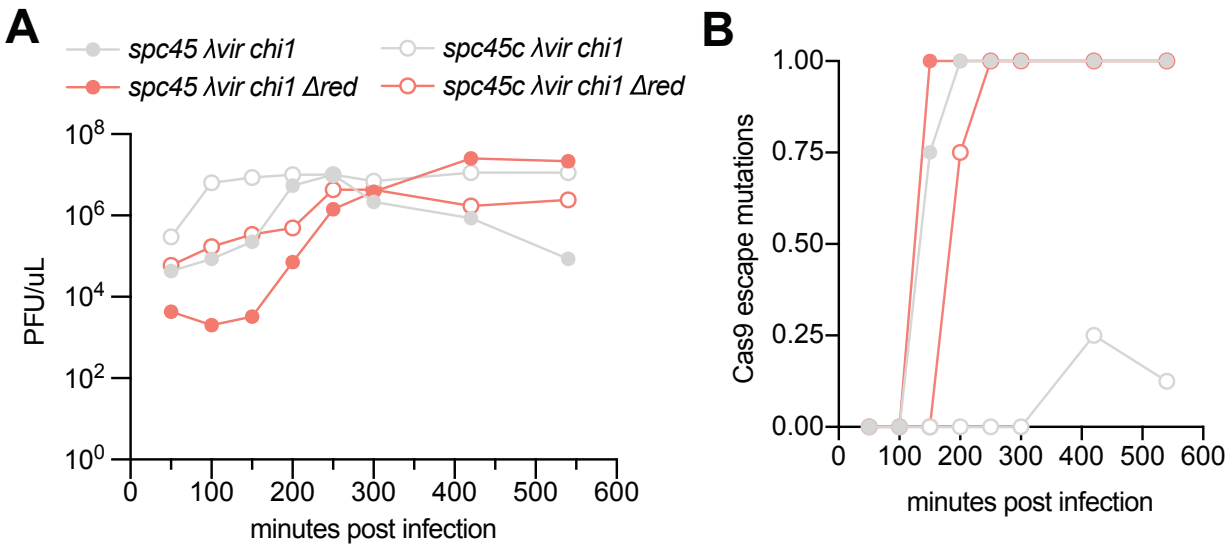


Figure 3.9 Inefficient Cas9 targeting allows the propagation of wild-type, unmutated phages that evade cleavage through λ Red recombination

(**A**) Titers of λ*vir chi*^{1} or λ*vir chi*^{1} Δ*red* at different times after infection of *E. coli* expressing Cas9 programmed with *spc45* or *spc45c* with λ*vir chi*^{1} or λ*vir chi*^{1} Δ*red* phages at an MOI of 1. (**B**) Fraction of plaque forming units harboring *spc45* target mutations from a total of 8 plaques analyzed per infection at each time point after infection of *E. coli* expressing Cas9 programmed with *spc45* or *spc45c* with λ*vir chi*^{1} or λ*vir chi*^{1} Δ*red* phages at an MOI of 1. PFU, plaque forming units.

Table 3.10 Sequences of CRISPR escape phages generated during a liquid culture infection time course of *spc45* targeting of λ*vir chi*[1] phages

time (mins post infection)	*E. coli* host	phage	target	plaque	protospacer	PAM
50	wild-type	λ*vir chi* [1]	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AAG
150	wild-type	λ*vir chi* [1]	*spc45*	escaper 4	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45*	escaper 5	ACTCGTCAGAATGAATATTATCAAGCAGAA	AGG
150	wild-type	λ*vir chi* [1]	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
150	wild-type	λ*vir chi* [1]	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AAG
150	wild-type	λ*vir chi* [1]	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGA
200	wild-type	λ*vir chi* [1]	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45*	escaper 1	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGA
200	wild-type	λ*vir chi* [1]	*spc45*	escaper 3	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
200	wild-type	λ*vir chi* [1]	*spc45*	escaper 5	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
200	wild-type	λ*vir chi* [1]	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
200	wild-type	λ*vir chi* [1]	*spc45*	escaper 8	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
250	wild-type	λ*vir chi* [1]	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
250	wild-type	λ*vir chi* [1]	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
250	wild-type	λ*vir chi* [1]	*spc45*	escaper 2	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
250	wild-type	λ*vir chi* [1]	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
250	wild-type	λ*vir chi* [1]	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGA
250	wild-type	λ*vir chi* [1]	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
250	wild-type	λ*vir chi* [1]	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
250	wild-type	λ*vir chi* [1]	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
250	wild-type	λ*vir chi* [1]	*spc45*	escaper 8	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG

time (mins post infection)	*E. coli* host	phage	target	plaque	protospacer	PAM
300	wild-type	λ*vir chi* [1]	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
300	wild-type	λ*vir chi* [1]	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
300	wild-type	λ*vir chi* [1]	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
300	wild-type	λ*vir chi* [1]	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
300	wild-type	λ*vir chi* [1]	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
300	wild-type	λ*vir chi* [1]	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGA
300	wild-type	λ*vir chi* [1]	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
300	wild-type	λ*vir chi* [1]	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
300	wild-type	λ*vir chi* [1]	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
420	wild-type	λ*vir chi* [1]	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
420	wild-type	λ*vir chi* [1]	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
420	wild-type	λ*vir chi* [1]	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
420	wild-type	λ*vir chi* [1]	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
420	wild-type	λ*vir chi* [1]	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
420	wild-type	λ*vir chi* [1]	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
420	wild-type	λ*vir chi* [1]	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
420	wild-type	λ*vir chi* [1]	*spc45*	escaper 7	ACTCGTCAGAATGAATATTATCAAGCTGCA	AGG
420	wild-type	λ*vir chi* [1]	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
540	wild-type	λ*vir chi* [1]	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
540	wild-type	λ*vir chi* [1]	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
540	wild-type	λ*vir chi* [1]	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
540	wild-type	λ*vir chi* [1]	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
540	wild-type	λ*vir chi* [1]	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
540	wild-type	λ*vir chi* [1]	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
540	wild-type	λ*vir chi* [1]	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
540	wild-type	λ*vir chi* [1]	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGA
540	wild-type	λ*vir chi* [1]	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG

Table 3.11 Sequences of CRISPR escape phages generated during a liquid culture infection time course of *spc45* targeting of λ*vir chi*[1] Δ*red* phages

time (mins post infection)	*E. coli* host	phage	target	plaque	protospacer	PAM
50	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	TTG
150	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 2	ACTCGTCAGAATGAATATTCTGAAGCAGCA	AAG
150	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 3	ACTCGTCAGAATGAATATTCTGAAGCAGCA	AAG
150	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
150	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
150	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
150	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
150	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
200	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
200	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
200	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
200	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
200	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
200	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
200	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGA
200	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
250	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
250	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
250	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
250	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
250	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
250	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
250	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
250	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
250	wild-type	λ*vir chi* [1] Δ*red*	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT

time (mins post infection)	*E. coli* host	phage	target	plaque	protospacer	PAM
300	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
300	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
300	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
300	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
300	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
300	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
300	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
300	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
300	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
420	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
420	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	TTG
420	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
420	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
420	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
420	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	TTG
420	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	TTG
420	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
420	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
540	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
540	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
540	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	TTG
540	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
540	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
540	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
540	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
540	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	ACG
540	wild-type	λ*vir chi* ^{1} *Δred*	*spc45*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	TTG

Table 3.12 Sequences of CRISPR escape phages generated during a liquid culture infection time course of *spc45c* targeting of λ*vir chi*[1] phages

time (mins post infection)	*E. coli* host	phage	target	plaque	protospacer	PAM
50	wild-type	λ*vir chi* [1]	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi* [1]	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi* [1]	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi* [1]	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi* [1]	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
250	wild-type	λ*vir chi* [1]	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
250	wild-type	λ*vir chi* [1]	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
250	wild-type	λ*vir chi* [1]	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
250	wild-type	λ*vir chi* [1]	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
250	wild-type	λ*vir chi* [1]	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
250	wild-type	λ*vir chi* [1]	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
250	wild-type	λ*vir chi* [1]	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
250	wild-type	λ*vir chi* [1]	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
250	wild-type	λ*vir chi* [1]	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG

time (mins post infection)	*E. coli* host	phage	target	plaque	protospacer	PAM
300	wild-type	λ*vir chi*^{1}	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
300	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
300	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
300	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
300	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
300	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
300	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
300	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
300	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
420	wild-type	λ*vir chi*^{1}	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
420	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
420	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 2	ACTCGTCAGAATGAATATTATCAAGCGGCA	AGG
420	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
420	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 4	ACTCGTCAGAATGAATATTATCAAGCAGAA	AGG
420	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
420	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
420	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
420	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
540	wild-type	λ*vir chi*^{1}	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
540	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
540	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
540	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
540	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
540	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
540	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
540	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 7	ACTCGTCAGAATTAATATTATCAAGCAGCA	AGG
540	wild-type	λ*vir chi*^{1}	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG

Table 3.13 Sequences of CRISPR escape phages generated during a liquid culture infection time course of *spc45c* targeting of λ*vir chi*^{1} Δ*red* phages

time (mins post infection)	*E. coli* host	phage	target	plaque	protospacer	PAM
50	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
50	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
100	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
150	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
200	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
200	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
200	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	TTG
200	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
200	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
200	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
250	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
250	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
250	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
250	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
250	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAATCAGCA	AGG
250	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
250	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
250	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
250	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAATCAGCA	AGG

time (mins post infection)	*E. coli* host	phage	target	plaque	protospacer	PAM
300	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
300	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
300	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGA
300	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
300	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
300	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
300	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
300	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
300	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
420	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
420	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCCGCA	AGG
420	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCACGCAGCA	AGG
420	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
420	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT
420	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
420	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGA
420	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 7	ACTCGTCAGGATGACTATTATCAAGCAGCA	AGG
420	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCTGCA	AGG
540	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	wild-type	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGG
540	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 1	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
540	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 2	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
540	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 3	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
540	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 4	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
540	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 5	ACTCGTCAGGATGAATATTATCAAGCAGCA	ATG
540	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 6	ACTCGTCAGGATGAATATTATCAAGCAGCA	TTG
540	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 7	ACTCGTCAGGATGAATATTATCAAGCAGAA	AGG
540	wild-type	λ*vir chi*^{1} Δ*red*	*spc45c*	escaper 8	ACTCGTCAGGATGAATATTATCAAGCAGCA	AGT

3.5 λ Red facilitates different types of Cas9 escape mutations

Next, we investigated if the Red system was also involved in the generation of the other types of escaper mutations (**Fig. 2.6**). For *spc15* targeting, absence of *exo* and *bet*, *gam* or the full deletion of the *red* operon reduced escaper formation by approximately one to two orders of magnitude (**Fig. 3.10**). Importantly, we detected qualitative differences in the target sequences of the escapers (**Table 3.14**). Similar results were obtained for *spc26D* (**Fig. 3.11** and **Table 3.15**). This spacer was also able to provide immunity in hosts expressing dCas9, however the number of escapers of dCas9 targeting was not reduced in the absence of the Red system (**Fig.**

3.11). This result corroborates a requirement for Cas9 target cleavage for the generation of escapers via Exo-Beta recombination of phage and chromosomal DNA.

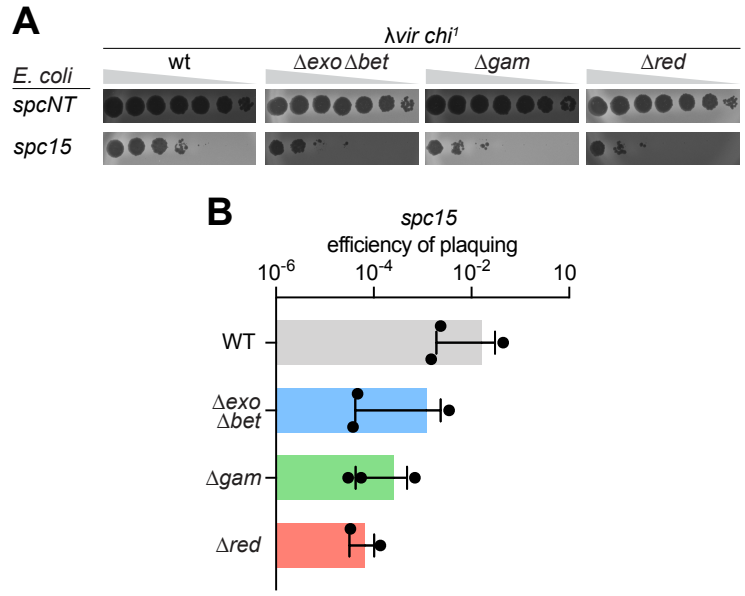


Figure 3.10 λ Red promotes the generation of host-recombined phage escapers

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*[1] wild-type and mutant phages containing deletions within the *red* genes on top agar plates seeded with *E. coli* expressing Cas9 programmed with *spc15*. *spcNT*, non-targeting spacer. (**B**) Efficiency of plaquing of different λ*vir chi*[1] phages on lawns of *E. coli* expressing Cas9 programmed with *spc15*. Mean ± SEM values of three independent experiments are shown.

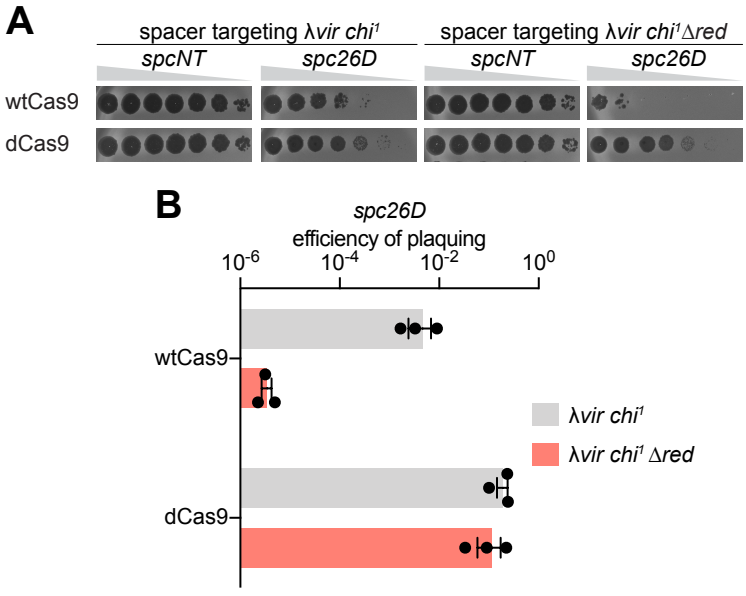


Figure 3.11 Cas9 cleavage is necessary for λ Red-mediated generation of host-recombined phage escapers

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*[1] or λ*vir chi*[1] Δ*red* phages on top agar plates seeded with *E. coli* expressing wild-type Cas9 or dCas9 programmed with *spc26D*. (**B**) Efficiency of plaquing of λ*vir chi*[1] or λ*vir chi*[1] Δ*red* phages on lawns of *E. coli* expressing Cas9 or dCas9 programmed with *spc26D*. Mean ± SEM values of three independent experiments are shown.

Table 3.14 Sequences of CRISPR escape phages generated during *spc15* targeting of λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages

E. coli host	phage	target	plaque	protospacer	PAM
wild-type	λ*vir chi* [1]	*spc15*	wild-type	CACCATCAGTTCAAGACGACGCAGCACCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 1	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 2	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 3	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 4	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 5	CACCATCAGTTCAAGACGACGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 6	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 7	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 8	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc15*	wild-type	CACCATCAGTTCAAGACGACGCAGCACCTC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc15*	escaper 1	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc15*	escaper 2	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc15*	escaper 3	CACCATCAGTTCAAGACGACGCAGCACCTC	CGT
wild-type	λ*vir chi* [1] Δ*red*	*spc15*	escaper 4	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc15*	escaper 5	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1] Δ*red*	*spc15*	escaper 6	CACCATCAGTTCAAGACGACGCAGCACCTC	CGT
wild-type	λ*vir chi* [1] Δ*red*	*spc15*	escaper 7	CACCATCAGTTCAAGACGACGCAGCACCTC	CTG
wild-type	λ*vir chi* [1] Δ*red*	*spc15*	escaper 8	CACCATCAGTTCAAGACGACGCAGCACCTC	CGT

Table 3.15 Sequences of CRISPR escape phages generated during *spc26D* targeting of λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages

E. coli host	phage	target	plaque	protospacer	PAM
wild-type	λ*vir chi* [1]	*spc26D*	wild-type	TACGGCGCAATTCCGCATCAGTAAGCGCAT	TGG
wild-type	λ*vir chi* [1]	*spc26D*	escaper 1	TGCGCCGTAATTCCGCGTCAGCCAGCGCAT	TGG
wild-type	λ*vir chi* [1]	*spc26D*	escaper 2	TGCGCCGTAATTCCGCGTCAGCCAGCGCAT	TGG
wild-type	λ*vir chi* [1]	*spc26D*	escaper 3	TGCGCCGTAATTCCGCGTCAGCCAGCGCAT	TGG
wild-type	λ*vir chi* [1]	*spc26D*	escaper 4	TGCGCCGTAATTCCGCGTCAGCCAGCGCAT	TGG
wild-type	λ*vir chi* [1]	*spc26D*	escaper 5	TGCGCCGTAATTCCGCGTCAGCCAGCGCAT	TGG
wild-type	λ*vir chi* [1]	*spc26D*	escaper 6	TGCGCCGTAATTCCGCGTCAGCCAGCGCAT	TGG
wild-type	λ*vir chi* [1]	*spc26D*	escaper 7	TGCGCCGTAATTCCGCGTCAGCCAGCGCAT	TGG
wild-type	λ*vir chi* [1]	*spc26D*	escaper 8	TGCGCCGTAATTCCGCGTCAGCCAGCGCAT	TGG
wild-type	λ*vir chi* [1] Δ*red*	*spc26D*	wild-type	TACGGCGCAATTCCGCATCAGTAAGCGCAT	TGG
wild-type	λ*vir chi* [1] Δ*red*	*spc26D*	escaper 1	TACGGCGCAATTCCGCATCAGTAAGCGCAT	TGT
wild-type	λ*vir chi* [1] Δ*red*	*spc26D*	escaper 2	TACGGCGCAATTCCGCATCAGTAAGCGCAT	TGT
wild-type	λ*vir chi* [1] Δ*red*	*spc26D*	escaper 3	TACGGCGCAATTCCGCATCAGTAAGCGCAT	TGT
wild-type	λ*vir chi* [1] Δ*red*	*spc26D*	escaper 4	TACGGCGCAATTCCGCATCAGTAAGCGCAT	TGT
wild-type	λ*vir chi* [1] Δ*red*	*spc26D*	escaper 5	TACGGCGCAATTCCGCATCAGTAAGCGCAT	TGT
wild-type	λ*vir chi* [1] Δ*red*	*spc26D*	escaper 6	TACGGCGCAATTCCGCATCAGTAAGCGCAT	TGT
wild-type	λ*vir chi* [1] Δ*red*	*spc26D*	escaper 7	TACGGCGCAATTCCGCATCAGTAAGCGCAT	TGT
wild-type	λ*vir chi* [1] Δ*red*	*spc26D*	escaper 8	TACGGCGCAATTCCGCATCAGTAAGCGCAT	TGT

For *spc14* targeting, deletion of *exo* and *bet*, *gam* or *red* decreased the number of plaques obtained by approximately two orders of magnitude (three orders of magnitude in the case of λ*vir chi*[1]Δ*red*) (**Fig. 3.12**). Importantly, when we analyzed the *spc14* target region of six escapers (**Table 3.16**), we found that while all λ*vir chi*[1] mutants contained the same microhomology-mediated target deletion, absence of Red recombination led to the accumulation of target point mutations in 5/6 of the λ*vir chi*[1]Δ*red* escapers.

The chromosomal recombination and microhomology-mediated deletions in the *spc15* and *spc14* targets respectively do not require the introduction of *de novo* point

mutations in the target during repair. Therefore, it is expected that Pol IV, involved in the generation of protospacer mutations during *spc9* targeting, should not affect the number of *spc15* and *spc14* escapers. To test this prediction, we performed infections in Δ*dinB* mutant hosts and found no difference in the number of *spc15* escapers and a mild reduction of *spc14* escapers (**Fig. 3.13**). In addition, sequencing of the escaper targets showed minimal differences in the type of mutations (**Table 3.17** and **Table 3.18**). Altogether these results demonstrate that Exo-Beta can also facilitate the repair of Cas9 DSBs through recombination with the host's chromosomal DNA and through MMEJ to generate escaper phages.

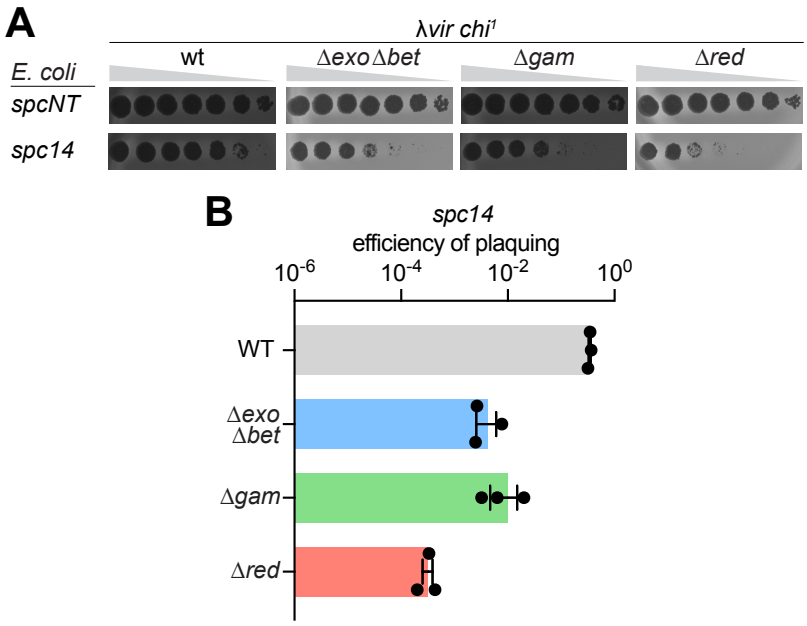


Figure 3.12 λ Red promotes the generation of microhomology-mediated target deletions

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*[1] wild-type and mutant phages containing deletions within the *red* genes on top agar plates seeded with *E. coli* expressing Cas9 programmed with *spc14*. *spcNT*, non-targeting spacer. (**B**) Efficiency of plaquing of different λ*vir chi*[1] phages on lawns of *E. coli* expressing Cas9 programmed with *spc14*. Mean ± SEM values of three independent experiments are shown.

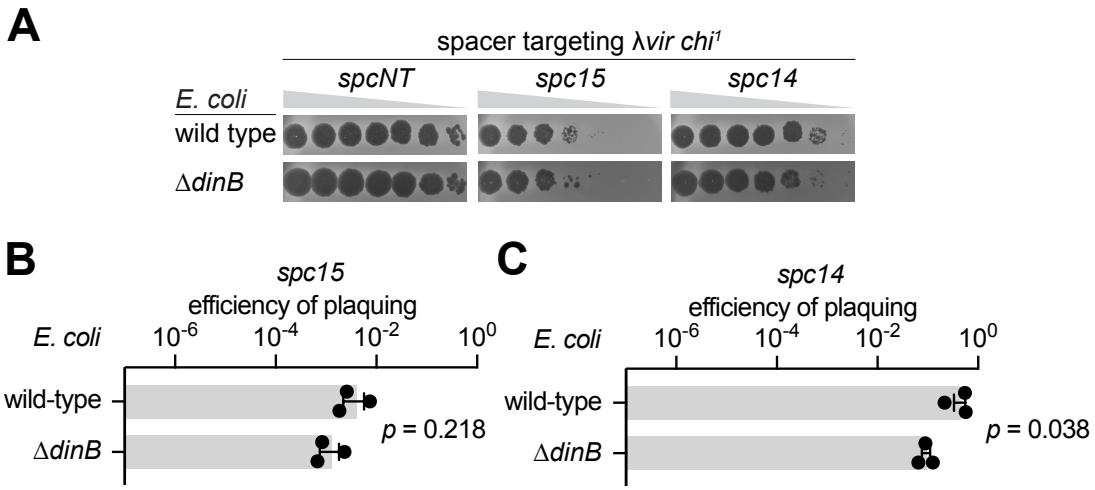


Figure 3.13 Pol IV is not involved in the generation of phage escapers arising through host chromosome recombination or microhomology-mediated end joining

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*[1] or λ*vir chi*[1] Δ*red* phages on top agar plates seeded with *E. coli* expressing Cas9 programmed with different spacers, in the presence or absence of *E. coli* Pol IV (*dinB*). (**B**-**C**) Efficiency of plaquing of λ*vir chi*[1] phage on lawns of *E. coli* expressing Cas9 programmed with *spc15* (**B**) or *spc14* (**C**), in the presence or absence of *E. coli* Pol IV DNA polymerase (*dinB*). Mean ± SEM values of three independent experiments are shown.

Table 3.16 Sequences of CRISPR escape phages generated during *spc14* targeting of λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages

***E. coli* host**	**phage**	**target**	**plaque**	**protospacer**	**PAM**
wild-type	λ*vir chi* [1]	*spc14*	wild-type	CAGGGGTGTTACCACTACCGCAGGAAAAGG	AGG
wild-type	λ*vir chi* [1]	*spc14*	escaper 1	398 bp deletion (MMEJ GAAGCTGCATG)	
wild-type	λ*vir chi* [1]	*spc14*	escaper 2	398 bp deletion (MMEJ GAAGCTGCATG)	
wild-type	λ*vir chi* [1]	*spc14*	escaper 3	398 bp deletion (MMEJ GAAGCTGCATG)	
wild-type	λ*vir chi* [1]	*spc14*	escaper 4	398 bp deletion (MMEJ GAAGCTGCATG)	
wild-type	λ*vir chi* [1]	*spc14*	escaper 5	398 bp deletion (MMEJ GAAGCTGCATG)	
wild-type	λ*vir chi* [1]	*spc14*	escaper 6	398 bp deletion (MMEJ GAAGCTGCATG)	
wild-type	λ*vir chi* [1] Δ*red*	*spc14*	wild-type	CAGGGGTGTTACCACTACCGCAGGAAAAGG	AGG
wild-type	λ*vir chi* [1] Δ*red*	*spc14*	escaper 1	398 bp deletion (MMEJ GAAGCTGCATG)	
wild-type	λ*vir chi* [1] Δ*red*	*spc14*	escaper 2	CAGGGGTGTTACCACTACCGCAGGAAAATG	AGG
wild-type	λ*vir chi* [1] Δ*red*	*spc14*	escaper 3	CAGGGGTGTTACCACTACCGCAGGAAAAGG	ATG
wild-type	λ*vir chi* [1] Δ*red*	*spc14*	escaper 4	CAGGGGTGTTACCACTACCGCAGGAAAATA	AGG
wild-type	λ*vir chi* [1] Δ*red*	*spc14*	escaper 5	CAGGGGTGTTACCACTACCGCAGGAAAATG	AGG
wild-type	λ*vir chi* [1] Δ*red*	*spc14*	escaper 6	CAGGGGTGTTACCACTACCGCAGGAAAAGG	ATG

Table 3.17 Sequences of CRISPR escape phages generated during *spc15* targeting of λ*vir chi*[1] in the presence or absence *E. coli* Pol IV (*dinB*)

E. coli host	phage	target	plaque	protospacer	PAM
wild-type	λ*vir chi* [1]	*spc15*	wild-type	CACCATCAGTTCAAGACGACGCAGCACCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 1	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 2	CACCATCAGTTCAAGACGACGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 3	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 4	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 5	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 6	CACCATCAGTTCAAGACGACGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 7	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
wild-type	λ*vir chi* [1]	*spc15*	escaper 8	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
Δ*dinB*	λ*vir chi* [1]	*spc15*	wild-type	CACCATCAGTTCAAGACGACGCAGCACCTC	CGG
Δ*dinB*	λ*vir chi* [1]	*spc15*	escaper 1	CACCATCAGTTCAAGACGACGCAGTGCCTC	CGG
Δ*dinB*	λ*vir chi* [1]	*spc15*	escaper 2	CACCATCAGTTCAAGACGACGCAGCACCTC	CGT
Δ*dinB*	λ*vir chi* [1]	*spc15*	escaper 3	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
Δ*dinB*	λ*vir chi* [1]	*spc15*	escaper 4	CACCATCAGTTCAAGACGGCGCAGTGCCTC	CGG
Δ*dinB*	λ*vir chi* [1]	*spc15*	escaper 5	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
Δ*dinB*	λ*vir chi* [1]	*spc15*	escaper 6	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
Δ*dinB*	λ*vir chi* [1]	*spc15*	escaper 7	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG
Δ*dinB*	λ*vir chi* [1]	*spc15*	escaper 8	CACCATCAGTTCAAAACGGCGCAGTGCCTC	CGG

Table 3.18 Sequences of CRISPR escape phages generated during *spc14* targeting of λ*vir chi*[1] in the presence or absence *E. coli* Pol IV (*dinB*)

E. coli host	phage	target	plaque	protospacer	PAM
wild-type	λ*vir chi* [1]	*spc14*	wild-type	CAGGGGTGTTACCACTACCGCAGGAAAAGG	AGG
wild-type	λ*vir chi* [1]	*spc14*	escaper 1	16 bp PAM side deletion (GAG flanks)	
wild-type	λ*vir chi* [1]	*spc14*	escaper 2	398 bp deletion (MMEJ GAAGCTGCATG)	
wild-type	λ*vir chi* [1]	*spc14*	escaper 3	1952 bp deletion (MMEJ TGTTCTG)	
wild-type	λ*vir chi* [1]	*spc14*	escaper 4	398 bp deletion (MMEJ GAAGCTGCATG)	
wild-type	λ*vir chi* [1]	*spc14*	escaper 5	1561 bp deletion (MMEJ TGCCGGGAATGG)	
wild-type	λ*vir chi* [1]	*spc14*	escaper 6	398 bp deletion (MMEJ GAAGCTGCATG)	
wild-type	λ*vir chi* [1]	*spc14*	escaper 7	1951 bp deletion (MMEJ TCGATACCGGC)	
wild-type	λ*vir chi* [1]	*spc14*	escaper 8	398 bp deletion (MMEJ GAAGCTGCATG)	
Δ*dinB*	λ*vir chi* [1]	*spc14*	wild-type	CAGGGGTGTTACCACTACCGCAGGAAAAGG	AGG
Δ*dinB*	λ*vir chi* [1]	*spc14*	escaper 1	788 bp deletion (MMEJ GCGAGG)	
Δ*dinB*	λ*vir chi* [1]	*spc14*	escaper 2	1755 bp deletion (MMEJ CGGGCAGG)	
Δ*dinB*	λ*vir chi* [1]	*spc14*	escaper 3	CAGGGGTGTTACCACTACCGCAGGAAA-GG	AGG
Δ*dinB*	λ*vir chi* [1]	*spc14*	escaper 4	843 bp deletion (MMEJ GCGAG)	
Δ*dinB*	λ*vir chi* [1]	*spc14*	escaper 5	2054 bp deletion (MMEJ GAATGCCTG)	
Δ*dinB*	λ*vir chi* [1]	*spc14*	escaper 6	1951 bp deletion (MMEJ TCGATACCGGC)	
Δ*dinB*	λ*vir chi* [1]	*spc14*	escaper 7	1952 bp deletion (MMEJ TGTTCTG)	
Δ*dinB*	λ*vir chi* [1]	*spc14*	escaper 8	1561 bp deletion (MMEJ TGCCGGGAATGG)	

3.6 The λ Red system enables escape from type I CRISPR-Cas targeting

To date, there are no natural isolates of *E. coli* that harbor a type II CRISPR-Cas system. Instead, this organism carries type I CRISPR loci (Touchon et al., 2011). In particular, the K-12 strain MG1655 harbors a type I-E CRISPR-Cas system that has been thoroughly characterized (Brouns et al., 2008; Semenova et al., 2011). This system is repressed at the transcription level under laboratory growth conditions (Westra et al., 2010), therefore we used the engineered strain ACT-01 in which the type I-E *cas* genes encoding Cascade and Cas3 are under the control of an arabinose-inducible promoter on the *E. coli* chromosome (Caliando and Voigt, 2015). *spc9R*, which targets phage λ in the same region specified by *spc9* (**Fig. 2.1-A**), was introduced under the control of the same promoter on a plasmid to provide crRNAs for the Cascade complex. Compared to the non-targeting control, type I-E targeting reduced the number of plaques by almost two orders of magnitude for the λ*vir chi*^{1} phage (**Fig. 3.14**). Absence of *exo* and *bet*, *gam* or *red* further reduced that number by approximately three to four orders of magnitude relative to the wild-type phage. Importantly, we found that while 7/8 *spc9R* λ*vir chi*^{1} targets contained a G>A substitution in position -4, only 1/8 λ*vir chi*^{1}Δ*red* escapers harbored that mutation (**Table 3.19**).

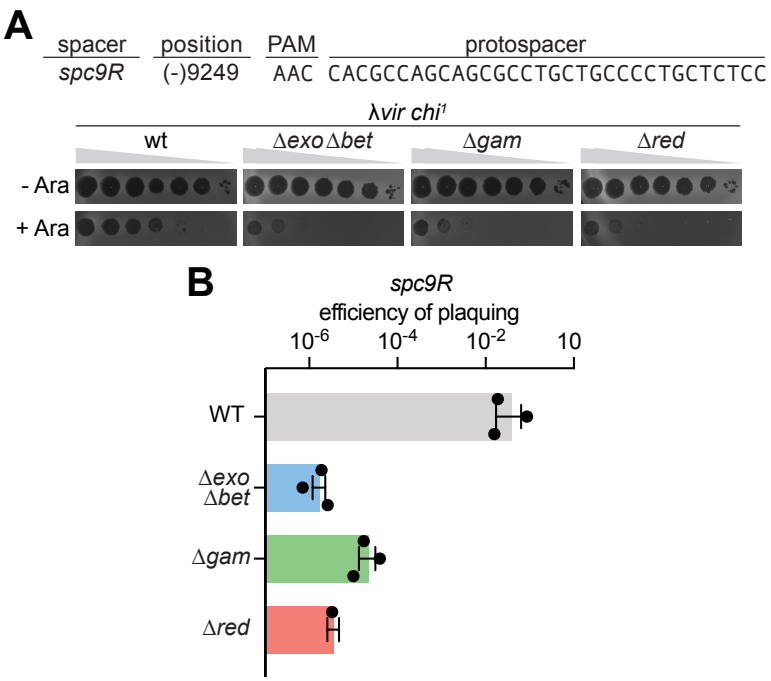


Figure 3.14 λ Red promotes escape from type I-E CRISPR-Cas targeting

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*[1] wild-type and mutant phages containing deletions within the *red* genes on top agar plates seeded with *E. coli* ACT-01 expressing the Cascade-Cas3 complex under the control of an arabinose-inducible promoter, programmed with *spc9R* on a plasmid with a separate arabinose-inducible promoter, in the presence or absence of the inducer. The genomic position and sequence of the *spc9R* target are shown. (**B**) Efficiency of plaquing of different λ*vir chi*[1] phages on lawns of *E. coli* ACT-01 expressing the Cascade-Cas3 complex programmed with *spc9R*. Mean ± SEM values of three independent experiments are shown.

Table 3.19 Sequences of CRISPR escape phages generated during type I-E *spc9R* targeting of λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages

E. coli host	phage	target	plaque	PAM	protospacer
ACT-01	λ*vir chi* [1]	Type I-E *spc9R*	wild-type	AAC	CACGCCAGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1]	Type I-E *spc9R*	escaper 1	AAC	CACACCAGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1]	Type I-E *spc9R*	escaper 2	AAC	CACACCAGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1]	Type I-E *spc9R*	escaper 3	AAC	CACACCAGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1]	Type I-E *spc9R*	escaper 4	AAC	CACACCAGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1]	Type I-E *spc9R*	escaper 5	AAC	CACACCAGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1]	Type I-E *spc9R*	escaper 6	AAC	CACACCAGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1]	Type I-E *spc9R*	escaper 7	AAC	CACGACAGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1]	Type I-E *spc9R*	escaper 8	AAC	CACACCAGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1] Δ*red*	Type I-E *spc9R*	wild-type	AAC	CACGCCAGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1] Δ*red*	Type I-E *spc9R*	escaper 1	AAC	CACGCCAGCATCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1] Δ*red*	Type I-E *spc9R*	escaper 2	AAC	CACGCCTGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1] Δ*red*	Type I-E *spc9R*	escaper 3	AAC	CACGCCATCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1] Δ*red*	Type I-E *spc9R*	escaper 4	AAC	CACGCCATCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1] Δ*red*	Type I-E *spc9R*	escaper 5	AAC	CACGCCTGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1] Δ*red*	Type I-E *spc9R*	escaper 6	AAC	CACACCAGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1] Δ*red*	Type I-E *spc9R*	escaper 7	AAC	CACTCCAGCAGCGCCTGCTGCCCCTGCTCTCC
ACT-01	λ*vir chi* [1] Δ*red*	Type I-E *spc9R*	escaper 8	AAC	CAAGCCAGCAGCGCCTGCTGCCCCTGCTCTCC

We also explored the effect of Red in the escape from the targeting provided by four "natural" spacers; i.e., that were acquired by the type I-E system of *E. coli* from large genomic fragments of λ DNA in a previous study, using strain KD263 (Strotskaya et al., 2017). Of these, only *spcE4-R* provided a mild defense comparable to *spc9R* (EOP ~ 10^{-2}, **Fig. 3.15-ABC**); the other three (*spcL1-R*, *spcL4-R* and *spcL6-R*) mediated a substantial decrease of λ*vir chi*^{1} EOP (**Fig. 3.15-ABDEF**). When cells carrying any of these spacers were infected with λ*vir chi*^{1}Δ*red*, the EOP values dropped at least two orders of magnitude compared to the values obtained for phages expressing the Red system (**Fig. 3.15**). As reported for *spc9R*, analysis of the target sequences of *spcL4-R* and *spcL6-R* escapers showed distinct mutations in the presence or absence of Red (**Table 3.20** and **Table 3.21**, respectively). This did not seem to be the case for *spcL1-R* and *spcE1-R* escapers, most of which contained small and complete target deletions, respectively (**Table 3.22** and data not shown). Finally, we investigated the role of the error-prone polymerases in the generation of escapers during *spcL6-R*-mediated Cascade-Cas3 targeting (**Fig. 3.16**). However, none of the DNA polymerase deletions affected the EOP values of λ*vir chi*^{1}. These observations demonstrate that the Red system also enables an increase in type I-E CRISPR escapers with a distinct mutational signature. However, as opposed to the evasion of type II-A targeting, Red facilitates the introduction of escape mutations when the immunity provided by the spacer is both weak and strong, and in the absence of Pol IV.

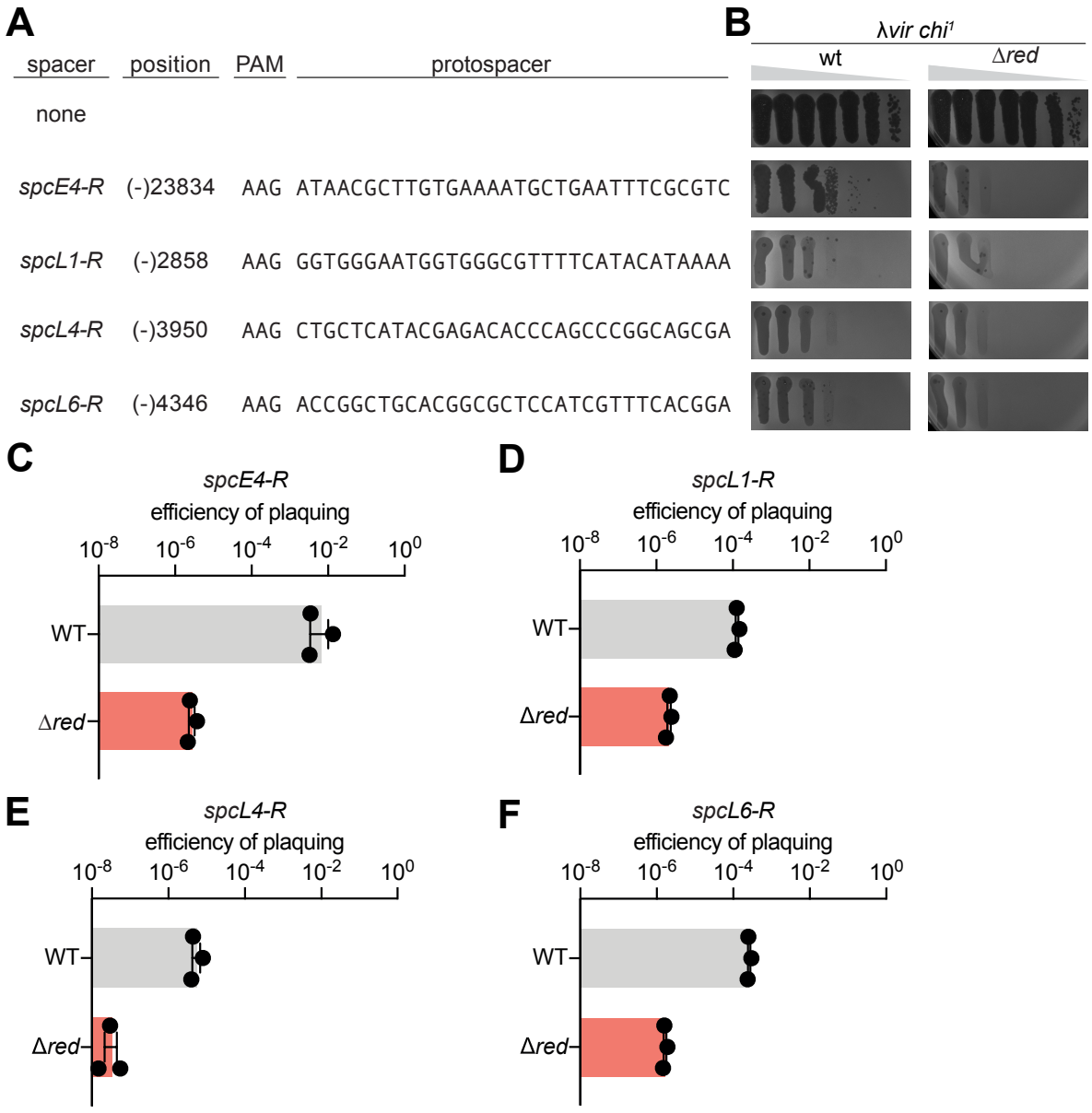


Figure 3.15 λ Red promotes escape from strong targeting type I-E CRISPR spacers

(**A**) Sequences of the type I-E targets specified by the anti-λ spacers acquired into strain KD263 that were used in this study. (**B**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*^{1} or λ*vir chi*^{1} Δ*red* phages on top agar plates seeded with *E. coli* KD263 and KD263-derived targeting strains carrying the spacers shown in (**A**); *cas* genes and spacers were induced on top agar with 1 mM arabinose and 1 mM IPTG. (**C**-**F**) Efficiency of plaquing of λ*vir chi*^{1} or λ*vir chi*^{1} Δ*red* phages on lawns of *E. coli* KD263 harboring *spcE4-R* (**C**), *spcL1-R* (**D**), *spcL4-R* (**E**), and *spcL6-R* (**F**). Mean ± SEM values of three independent experiments are shown.

Table 3.20 Sequences of CRISPR escape phages generated during type I-E *spcL4-R* targeting of λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages

E. coli host	phage	target	plaque	PAM	protospacer	3' end
λ_*L4-R*	λ*vir chi* [1]	Type I-E *spcL4-R*	wild-type	AAG	CTGCTCATACGAGACACCCAGCCCGGCAGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1]	Type I-E *spcL4-R*	escaper 1	AAT	TTGCTCATACGAGACACCCAGCCCGGCAGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1]	Type I-E *spcL4-R*	escaper 2	GAT	CTGCTCATACGAGACACCCAGCCCGGCAGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1]	Type I-E *spcL4-R*	escaper 3	AAG	TTGTTCATACGAGACACCCAGCCCGGCAGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1]	Type I-E *spcL4-R*	escaper 4	AAT	TTGCTCATACGAGACACCCAGCCCGGCAGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1]	Type I-E *spcL4-R*	escaper 5	CAG	CTGTTCATACGAGACACCCAGCCCGGCAGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1]	Type I-E *spcL4-R*	escaper 6	AAT	CTGCTCATACGAGACACCCAGCCCGGCAGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1]	Type I-E *spcL4-R*	escaper 7	AAG	TTGTTCATACGAGACACCCAGCCCGGCAGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1]	Type I-E *spcL4-R*	escaper 8	AAT	TTGCTCATACGAGACACCCAGCCCGGCAGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL4-R*	wild-type	AAG	CTGCTCATACGAGACACCCAGCCCGGCAGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL4-R*	escaper 1	AAG	CTGCTCATACGAGACACCCAGCCCTGCCGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL4-R*	escaper 2	AAG	CTGCTCATACGAGACACCCAGCCCGGCAGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL4-R*	escaper 3	AAG	CTGCTCATACGAGACACCCAGCCCTGCCGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL4-R*	escaper 4	AAG	CTGCTCATACGAGACACCCAGCCCTGCTGCGG	TATAC
λ_*L4-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL4-R*	escaper 5	AAG	CTGCTCATACGAGACACCCAGCCCTGCTGCGG	TATAT
λ_*L4-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL4-R*	escaper 6	AAG	CTGCTCATACGAGACACCCAGCCCTGCCGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL4-R*	escaper 7	AAG	CTGCTCATACGAGACACCCAGCCCTGCCGCGA	TATAC
λ_*L4-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL4-R*	escaper 8	AAG	CTGCTCATACGAGACACCCATCCCTGCCGCGA	TATAC

Table 3.21 Sequences of CRISPR escape phages generated during type I-E *spcL6-R* targeting of λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages

E. coli host	phage	target	plaque	PAM	protospacer
λ_*L6-R*	λ*vir chi* [1]	Type I-E *spcL6-R*	wild-type	AAG	ACCGGCTGCACGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1]	Type I-E *spcL6-R*	escaper 1	AAG	GCCGGCGGAACGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1]	Type I-E *spcL6-R*	escaper 2	AAG	GTCGGCTGCACGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1]	Type I-E *spcL6-R*	escaper 3	AAG	GCCAGCTGCACGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1]	Type I-E *spcL6-R*	escaper 4	AAG	GCCGGCCGCACGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1]	Type I-E *spcL6-R*	escaper 5	AAG	GCCGGCGGCACGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1]	Type I-E *spcL6-R*	escaper 6	AAG	GCCGGCAGCCCGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1]	Type I-E *spcL6-R*	escaper 7	AAG	GCCGGCCGCACGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1]	Type I-E *spcL6-R*	escaper 8	GAG	GCCGGCTGCACGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL6-R*	wild-type	AAG	ACCGGCTGCACGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL6-R*	escaper 1	AAG	TCCGGCGGCACGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL6-R*	escaper 2	AAG	CCCGGCTGCCCGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL6-R*	escaper 3	AAG	ACCGGCGATCCGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL6-R*	escaper 4	CAG	CCCGGCTGCACGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL6-R*	escaper 5	CAG	ACCGGCTGCACGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL6-R*	escaper 6	AAG	ACCGGCGGTGCGGCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL6-R*	escaper 7	AAG	ACCGGCCAAACTCCGCTCCATCGTTTCACGGA
λ_*L6-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL6-R*	escaper 8	AAG	TCCGGCGGCACGGCGCTCCATCGTTTCACGGA

Table 3.22 Sequences of CRISPR escape phages generated during type I-E *spcL1-R* targeting of λ*vir chi*[1] and λ*vir chi*[1] Δ*red* phages

E. coli host	phage	target	plaque	PAM	protospacer
λ_*L1-R*	λ*vir chi* [1]	Type I-E *spcL1-R*	wild-type	AAG	GGTGGGAATGGTGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1]	Type I-E *spcL1-R*	escaper 1	AAG	GGGGGGGGTGGTGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1]	Type I-E *spcL1-R*	escaper 2	AAG	GGTGGGAATGGTGGGCG-TTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1]	Type I-E *spcL1-R*	escaper 3	AAG	GGTGGGAATGGTGGGCG-TTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1]	Type I-E *spcL1-R*	escaper 4	AAG	---------GGTGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1]	Type I-E *spcL1-R*	escaper 5	AAG	---------GGTGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1]	Type I-E *spcL1-R*	escaper 6	AAG	---------GGTGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1]	Type I-E *spcL1-R*	escaper 7	AAG	GGTGGGAATGGTGGGCG-TTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1]	Type I-E *spcL1-R*	escaper 8	AAG	---------GGTGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL1-R*	wild-type	AAG	GGTGGGAATGGTGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL1-R*	escaper 1	AAG	---------GGTGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL1-R*	escaper 2	AAG	---------GGTGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL1-R*	escaper 3	AAG	---------GGTGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL1-R*	escaper 4	AAG	---------GGTGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL1-R*	escaper 5	AAG	GGTGGGAATGGTGGGCGTTTT-ATACATAAAA
λ_*L1-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL1-R*	escaper 6	AAG	GGTGGGAA---TGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL1-R*	escaper 7	AAG	---------GGTGGGCGTTTTCATACATAAAA
λ_*L1-R*	λ*vir chi* [1] Δ*red*	Type I-E *spcL1-R*	escaper 8	AAG	---------GGTGGGCGTTTTCATACATAAAA

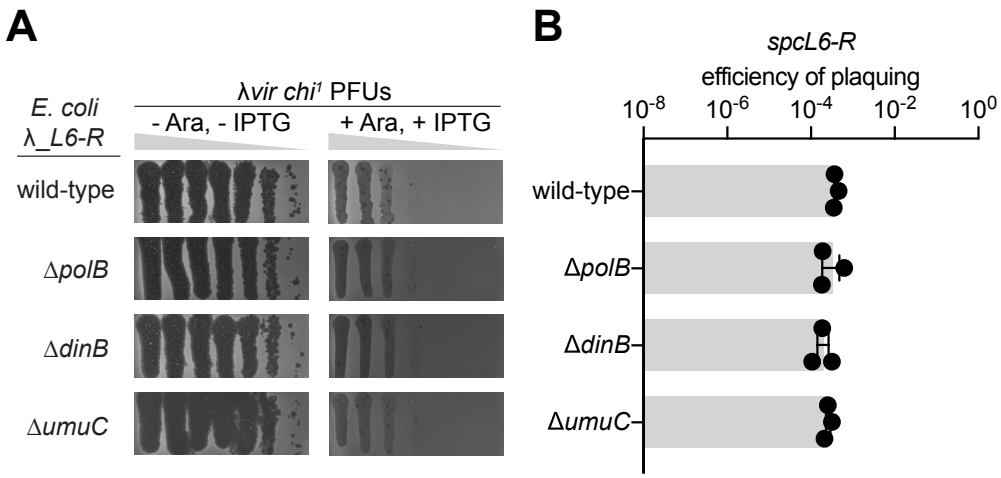


Figure 3.16 *E. coli* error-prone DNA polymerases do not promote phage escape from type I-E CRISPR-Cas targeting

(**A**) Detection of plaque formation after spotting 10-fold serial dilutions of λ*vir chi*[1] phage on top agar plates seeded with *E. coli* KD263 expressing type I-E spacer *L6-R*, in the presence or absence of genes encoding different *E. coli* error-prone DNA polymerases; *cas* genes and spacers were either uninduced or induced on top agar with 1 mM arabinose and 1 mM IPTG. (**B**) Efficiency of plaquing of λ*vir chi*[1] phage on lawns of *E. coli* KD263 harboring *spcL6-R*, in the presence or absence of genes encoding different *E. coli* error-prone DNA polymerases. Mean ± SEM values of three independent experiments are shown.

3.7 Summary

In Chapters 2 and 3, we investigated whether bacteriophage-encoded recombination systems can mediate viral escape from bacterial CRISPR-Cas immune systems. To address our overarching question, we used the well-described Red recombination system present in the extensively studied laboratory coliphage, phage λ. Using this model system, we addressed the following questions: (i) Does bacteriophage recombination mediate evasion of CRISPR-Cas targeting? (ii) Does bacteriophage recombination promote the generation of CRISPR-Cas escape mutations? These questions were addressed in Chapters 2 and 3, respectively.

In Chapter 2, we showed that λ Red mediates evasion of type II CRISPR-Cas targeting and that all three genes within the λ *red* operon, encoding Gam, Exo and Beta proteins, are necessary for immune evasion. Specifically, Gam inhibits the host-encoded recombinase RecBCD, which would otherwise degrade phage DNA following Cas9 cleavage. Exo and Beta can then act on the free DNA ends to mediate recombination and repair of the Cas9-cleaved phage DNA. Interestingly, we found that the phage-encoded λ Red system is more efficient than the host's RecBCD-RecA recombination system in promoting phage escape from Cas9 targeting, even when the host system is rigged to undergo constitutive recombination at free DNA ends.

In Chapter 3, we demonstrated that λ Red promotes the generation of target mutations during Cas9 targeting, which facilitate genetic escape from type II CRISPR-Cas targeting. Furthermore, we explored the range of target mutations enabled through λ Red recombination, including point mutations, microhomology-mediated deletions, and recombination with homoeologous cryptic prophage sequences in the bacterial chromosome. Finally, we showed that λ Red enables evasion of not only type II but also type I CRISPR-Cas targeting, using the type I-E CRISPR-Cas immune system native to *E. coli*. Altogether, we conclude that the phage-encoded Red recombination system effectively limits CRISPR-Cas targeting and promotes the generation of escape mutations.

CHAPTER 4. METAGENOMIC SCREENING IDENTIFIES A PROKARYOTIC DNA GLYCOSYLASE INVOLVED IN ANTIVIRAL DEFENSE

4.1 Background

In Chapters 2 and 3, I presented results that demonstrate a role for bacteriophage-encoded DNA repair systems in the arms race between bacteria and their viruses, specifically in providing a survival advantage to bacteriophages in the face of CRISPR attack. Now, I will switch focus and pivot to the flip side of the arms race, i.e., by examining an example of how bacteria can co-opt DNA repair proteins as part of their immune repertoire to fight off phages. Indeed, bacteria have evolved a diverse battery of immune systems to counteract infections by viruses and plasmids (Bernheim and Sorek, 2020; Rostol and Marraffini, 2019a). These prokaryotic immune systems typically cluster together in genetic loci termed “defense islands” (Makarova et al., 2013; Makarova et al., 2011). Multiple studies have used bioinformatic analysis to identify new antiviral genes and immune systems present within these islands (Doron et al., 2018; Gao et al., 2020; Johnson et al., 2022; Millman et al., 2022) as well as in pathogenicity islands (Fillol-Salom et al., 2022) and prophage elements (Rousset et al., 2022). More recently, one study performed a functional selection for antiviral immune systems in the *E. coli* pangenome, identifying immune systems harbored by *E. coli* strains that had gone unnoticed from previous bioinformatic searches (Vassallo et al., 2022). While these studies have undoubtedly expanded our understanding of the diversity of immune systems present in bacteria,

they have all relied on the availability of sequenced genomes. Analyses of 16S ribosomal RNA sequences suggest that uncultivated and unsequenced microbes represent the majority of bacterial lineages (Rappe and Giovannoni, 2003). This “microbial dark matter”, which is beginning to be accessed through single-cell genomics (Marcy et al., 2007; Rinke et al., 2013), arguably contains a vast assortment of unknown genetic pathways, including those involved in anti-phage defense. To tap into this uncharted sequence space, we screened a library of environmental DNA (eDNA) constructed in *E. coli* for clones showing resistance or immunity to coliphage T4 infection. This library was a generous gift from Dr. Sean Brady’s laboratory at the Rockefeller University. The library was constructed by cloning microbial DNA isolated from an arid soil collected in Arizona into a cosmid vector (Brady, 2007; Feng et al., 2011). We picked soil as the source of genetic material due to its high microbial presence and diversity (Bardgett and van der Putten, 2014; Ramirez et al., 2014), known to contain thousands of bacterial phylotypes per sample (Delgado-Baquerizo et al., 2018). We chose to challenge this library with the lytic coliphage T4 due to its well characterized molecular biology (Karam and Miller, 2010; Miller et al., 2003). Using this approach, we isolated Brig1, a DNA glycosylase from an unknown bacterium that provides immunity to bacteriophage T4 through the removal of alpha-glucosyl-hydroxymethylcytosine nucleobases present in the viral DNA (Lehman and Pratt, 1960). Brig1 generates abasic sites throughout the phage genome, which inhibit phage DNA replication upon infection. Since phages modify DNA bases in their genomes to counteract cleavage

by restriction and CRISPR-Cas nucleases (Bryson et al., 2015; Hutinet et al., 2019; Liu et al., 2020b; Luria and Human, 1952; Revel, 1967; Vlot et al., 2018), Brig1 most likely represents an evolutionary development of the host to regain immunity against phages that glucosylate their nucleobases. Our study illustrates both a powerful method for the discovery of new anti-phage defense systems, described in this chapter, as well as a unique mechanism of anti-phage immunity, introduced here and described further in subsequent chapters.

4.2 Infection of *E. coli* harboring an eDNA library uncovers a gene that protects against phage T4

To uncover novel anti-phage defense systems present in unsequenced bacterial genomes, we screened an eDNA library generated in an earlier study, harboring large (~40 kb) DNA fragments extracted from a soil sample collected in Arizona that were cloned into pWEB-TNC cosmids, packaged into λ phage and transfected into *E. coli* EC100 cells (Brady, 2007; Feng et al., 2011). The library screening was done in collaboration with Christian Baca, a talented graduate student in the Marraffini lab, who I was mentoring during his lab rotation at the time. Together, we infected the eDNA library, containing at least 10 million clones, with phage T4, a thoroughly characterized lytic coliphage (Karam and Miller, 2010; Miller et al., 2003), at a multiplicity of infection (MOI) of 10, to select resistant clones that can form a colony. To eliminate bacteria harboring chromosomal mutations that enable survival, we extracted cosmid DNA from surviving cells, re-transformed the

cosmids into fresh *E. coli* EC100 competent cells, and re-infected this new library with T4 at MOI 10 (**Fig. 4.1**). We repeated this process two times to enrich for *bona fide* resistant clones, after which we picked sixteen random colonies and tested their immunity using a plaque assay. We found that twelve of the sixteen colonies carried immunity to T4 phage infection, but did not provide immunity to an unrelated phage, λ*vir* (**Fig. 4.2**).

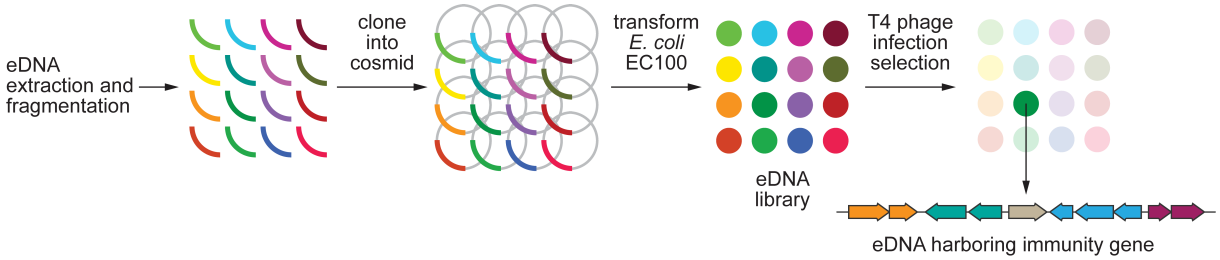


Figure 4.1 Selection of soil metagenomic DNA fragments that provide immunity against phage T4 in *E. coli*

Schematic showing the experimental setup for the screening of soil DNA (environmental DNA, eDNA) libraries for the presence of clones resistant to T4 infection. eDNA is extracted from soil samples and cleaved into large fragments of ~40 kb that are cloned into *E. coli* EC100. The library is infected with phage T4 in top agar plates to isolate surviving colonies. The cosmid from the T4-resistant colonies is extracted, sequenced, and further analyzed to identify and confirm immunity genes.

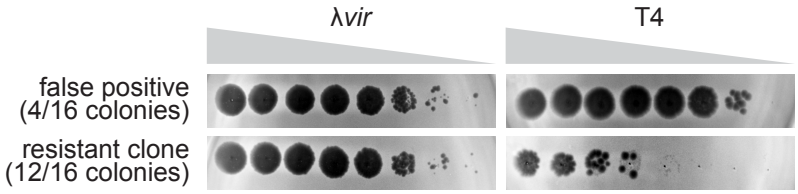


Figure 4.2 Identification of metagenomic library clones that provide immunity against phage T4 in *E. coli*

Ten-fold serial dilutions of λ*vir* or T4 phage on lawns of *E. coli* EC100 cells generated from one of 16 colonies sampled randomly from the eDNA library population that survived T4 infection. Top row is a representative result for a surviving colony that did not contain a *bona fide* immunity gene in the eDNA (4/16 false positives). Bottom row is a representative result for a surviving colony that harbored a cosmid carrying immunity genes (12/16 true positives).

Sequencing of the T4 phage-selected cosmids revealed that they all came from a single library clone, a 34.5 kb DNA insert of bacterial origin (**Fig. 4.3-A**), likely belonging to the phylum Actinobacteria, with homology to sequences present in the genera *Nocardioides*, *Gordonia* or *Mycobacterium* (**Table 4.1** and **Table 4.2**). To identify the genes responsible for T4 immunity, we divided the fragment into four different subclones, A-D, and found that subclone D retained the selected phenotype (**Fig. 4.3-AB**). Further subdivisions determined that fragment D3 harbored the putative immunity genes (**Fig. 4.3-CD**). This fragment contained a three-gene operon, gene *a* without homology, gene *b* with predicted ADP-ribosyl glycohydrolase activity and gene *c* encoding an unknown protein belonging to the superfamily of uracil DNA glycosylases (**Fig. 4.3-C**). Genetic dissection of the operon demonstrated that the latter was solely responsible for the immunity of *E. coli* hosts to T4 infection (**Fig. 4.4-A**). Moreover, expression of gene *c* using an arabinose-inducible promoter showed immunity only in the presence of the transcription inducer (**Fig. 4.4-B**), a result which demonstrates that this gene drives the observed T4 immunity.

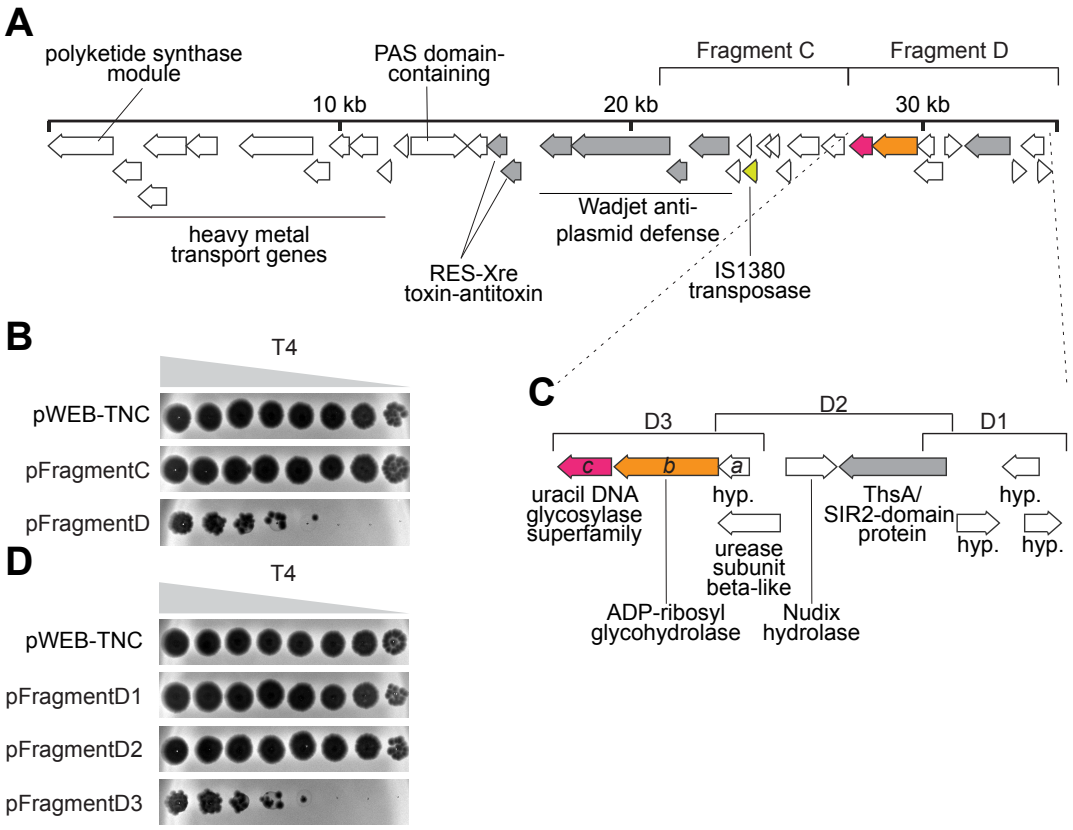


Figure 4.3 Subcloning of the metagenomic DNA fragment that provided immunity against phage T4

(**A**) Genes present within the 34.5 kb soil metagenomic DNA fragment that provided T4 immunity. Subcloned regions (Fragments C and D) are indicated. Transposases and known phage defense genes are shown in yellow and grey, respectively. (**B**) Ten-fold serial dilutions of phage T4 on lawns of *E. coli* EC100 carrying pWEB-TNC or cosmids containing Fragments C or D. (**C**) Genes present within Fragment D, showing the subclones generated for further analysis, Fragments D1-3; hyp, hypothetical protein. (**D**) Ten-fold serial dilutions of phage T4 on lawns of *E. coli* EC100 carrying pWEB-TNC or cosmids containing Fragments D1-3. Plaque images of one representative experiment from $n = 2$ independent experiments are shown in (**B**) and (**D**).

Table 4.1 Top 50 nucleotide BLAST hits of the Fragment C DNA sequence

#	Organism	Query Cover	% identity
1	*Gordonia iterans*	36%	75.75
2	*Mycobacterium ostraviense*	35%	73.88
3	*Mycolicibacterium rutilum*	36%	73.78
4	*Gordonia sp. 1D*	35%	74.11
5	*Gordonia amicalis*	35%	74.11
6	*Gordonia ajococcus*	35%	74.07
7	*Gordonia amicalis*	35%	74.07
8	*Gordonia amicalis*	35%	73.98
9	*Gordonia alkanivorans*	35%	73.67
10	*Gordonia sp. 135*	35%	73.63
11	*Gordonia alkanivorans*	35%	73.63
12	*Mycolicibacterium austroafricanum*	38%	73.13
13	*Mycolicibacterium vanbaalenii PYR-1*	36%	73.13
14	*Mycolicibacterium austroafricanum*	40%	73.09
15	*Mycolicibacterium austroafricanum*	40%	73.09
16	*Mycolicibacterium austroafricanum*	36%	73.29
17	*Mycobacterium sp. SMC-8*	39%	72.99
18	*Mycobacterium canettii CIPT 140070017*	35%	73.3
19	*Mycobacterium kansasii*	35%	73.2
20	*Mycobacterium kansasii*	35%	73.2
21	*Mycobacterium kansasii*	35%	73.2
22	*Mycobacterium kansasii*	35%	73.2
23	*Mycobacterium kansasii*	35%	73.2
24	*Mycobacterium kansasii*	35%	73.2
25	*Mycobacterium kansasii*	35%	73.2
26	*Mycobacterium kansasii*	35%	73.2
27	*Mycobacterium kansasii*	35%	73.2
28	*Mycobacterium kansasii*	35%	73.2
29	*Mycobacterium kansasii ATCC 12478*	35%	73.2
30	*Mycobacterium riyadhense*	35%	73.25
31	*Mycobacterium riyadhense*	35%	73.25
32	*Mycobacterium kansasii*	35%	73.24
33	*Mycobacterium kansasii*	35%	73.16
34	*Mycobacterium kansasii*	35%	73.16
35	*Mycolicibacterium smegmatis*	37%	73.16
36	*Mycolicibacterium smegmatis*	37%	73.16
37	*Mycolicibacterium smegmatis*	37%	73.16
38	*Mycolicibacterium smegmatis MKD8*	39%	73.11
39	*Mycolicibacterium smegmatis*	37%	73.11
40	*Mycolicibacterium smegmatis*	37%	73.11
41	*Mycolicibacterium smegmatis*	37%	73.11
42	*Mycolicibacterium smegmatis*	37%	73.03
43	*Mycolicibacterium smegmatis*	37%	73.03
44	*Mycolicibacterium smegmatis*	37%	73.03
45	*Mycolicibacterium smegmatis*	37%	73.03
46	*Mycolicibacterium smegmatis*	37%	73.03
47	*Mycolicibacterium smegmatis*	37%	73.03
48	*Mycolicibacterium smegmatis*	37%	73.03
49	*Mycolicibacterium smegmatis*	37%	73.03
50	*Mycolicibacterium smegmatis*	37%	73.03

Table 4.2 Top 50 nucleotide BLAST hits of the Fragment D DNA sequence

#	Organism	Query Cover	% identity
1	*Nocardioides coralli*	12%	72.48
2	*Nocardioides cynanchi*	12%	71.89
3	*Nocardioides sp. zg-536*	20%	68.51
4	*Nocardioides sp. S5*	8%	76.11
5	*Nocardioides ochotonae*	14%	71.24
6	*Nocardioides yefusunii*	14%	70.18
7	*Nocardioides sp. JQ2195*	18%	67.49
8	*Nocardioides euryhalodurans*	19%	69.85
9	*Miniimonas sp. S16*	8%	74.51
10	*Nocardioides baekrokdamisoli*	12%	67.87
11	*Microlunatus sagamiharensis*	13%	67.88
12	*Mycolicibacterium arabiense*	12%	67.36
13	*Pimelobacter simplex*	17%	67.16
14	*Pimelobacter simplex*	17%	67.16
15	*Nocardioides aromaticivorans*	12%	66.85
16	*Friedmanniella luteola*	16%	66.7
17	*Dietzia kunjamensis*	11%	66.44
18	*Mycobacterium grossiae*	12%	66.14
19	*Gordonia sp. JH63*	12%	66.67
20	*Nocardioides aquaticus*	19%	66.47
21	*Gordonia pseudamarae*	7%	70.76
22	*Gordonia pseudamarae*	7%	70.76
23	*Dietzia sp. B32*	11%	66.82
24	*Gordonia amarae*	8%	69.9
25	*Gordonia amarae*	8%	69.9
26	*Gordonia amarae*	8%	69.9
27	*Nocardioides sambongensis*	19%	65.67
28	*Janibacter sp. YB324*	25%	70.29
29	*Mycolicibacter sinensis*	7%	68.94
30	*Gordonia mangrovi*	14%	66.48
31	*Rhodococcus pyridinivorans*	13%	65.94
32	*Nocardioides sp. S-1144*	9%	73.33
33	*Ornithinimicrobium flavum*	11%	66.4
34	*Rhodococcus pyridinivorans*	13%	65.62
35	*Rhodococcus sp. Z13*	13%	65.4
36	*Rhodococcus rhodochrous*	13%	65.87
37	*Rhodococcus rhodochrous*	13%	65.87
38	*Rhodococcus pyridinivorans*	13%	65.51
39	*Rhodococcus sp. GA1*	13%	65.73
40	*Rhodococcus pyridinivorans*	13%	65.73
41	*Rhodococcus pyridinivorans*	13%	65.73
42	*Rhodococcus pyridinivorans SB3094*	13%	65.73
43	*Janibacter indicus*	17%	66.18
44	*Rhodococcus sp. 2G*	13%	65.41
45	*Rhodococcus pyridinivorans*	13%	65.41
46	*Rhodococcus pyridinivorans*	13%	65.41
47	*Rhodococcus pyridinivorans*	13%	65.41
48	*Rhodococcus sp. p52*	13%	65.38
49	*Rhodococcus pyridinivorans*	13%	65.62
50	*Rhodococcus pyridinivorans*	13%	65.62

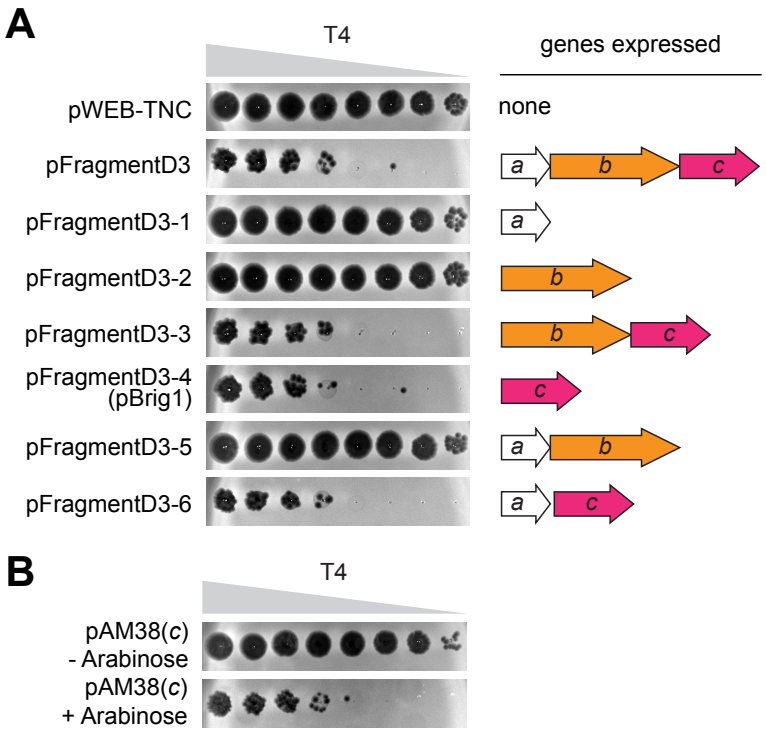


Figure 4.4 Isolation of the gene that provides immunity against phage T4

(**A**) Ten-fold serial dilutions of phage T4 on lawns of *E. coli* EC100 that harbor the pWEB-TNC cosmid carrying different genes present in a three-gene operon isolated after the screening of an eDNA library with this phage. Gene *c* encodes Brig1. (**B**) Same as (**A**) but infecting lawns of bacteria expressing Brig1 using an arabinose-inducible promoter, in the presence (+ Arabinose) or absence (- Arabinose) of the inducer. Plaque images of one representative experiment from $n = 3$ independent experiments are shown in (**A**) and (**B**).

Importantly, gene *c* is adjacent to other putative phage defense systems (**Fig. 4.3-AC**). These include a Thoeris A (ThsA)-like protein containing a SIR2 domain, commonly present in defense-associated sirtuins (DSRs) (Garb et al., 2022). ThsA in Thoeris defense systems hydrolyzes NAD+ to kill infected hosts upon detection of phage infection (Leavitt et al., 2022; Ofir et al., 2021). Also within the vicinity of gene *c* are a Wadjet anti-plasmid operon (Deep et al., 2022; Liu et al., 2022) and a RES-

Xre toxin-antitoxin (TA) pair, which could potentially be involved in anti-phage defense (LeRoux and Laub, 2022). This genetic context suggests that gene *c* is part of a bacterial defense island present in the isolated eDNA.

4.3 The prokaryotic DNA glycosylase Brig1 prevents T4 phage replication

To determine how gene *c* affects T4 infection, we first tested its impact on phage adsorption. Ten minutes after infection with T4, phage titers were equally reduced in cultures expressing gene *c* or harboring an empty vector, a result that indicates that the defense mechanism does not affect the rate of T4 adsorption (**Fig. 4.5**). At later timepoints, however, T4 plaque-forming units (PFUs) increased in control cells containing the empty vector, but decreased in the presence of gene *c*. To determine if disruption of T4 DNA replication was responsible for the lack of viral propagation, we performed quantitative PCR (qPCR) to measure phage DNA accumulation [at two different loci, *gp43* (encoding T4 DNA polymerase) and *gp34* (encoding the proximal unit of the long tail fiber)] in infected cells at 2-, 4-, 8- and 20-minutes post infection (**Fig. 4.6**). In contrast to susceptible hosts, which showed a steady increase in phage DNA over time, the T4 DNA content decreased in *E. coli* expressing gene *c* (**Fig. 4.6**). This result demonstrates that gene *c* not only inhibits T4 DNA replication but also causes a slight and gradual depletion of the phage DNA within the infected population (**Fig. 4.6-CD**). Finally, we performed next-generation sequencing (NGS) of *E. coli* cells infected with T4 for 8 minutes, which showed that

phage DNA reads were severely depleted across the entire T4 genome in gene *c*-expressing cells relative to non-immune control cells (**Fig. 4.7**). Altogether, these experiments demonstrate that the putative DNA glycosylase encoded by gene *c* prevents the replication of T4 viral DNA, and therefore we named this gene bacteriophage replication inhibition DNA glycosylase 1, Brig1. The cosmid harboring only gene *c*, pFragmentD3-4 (**Fig. 4.4-A**), was therefore renamed pBrig1.

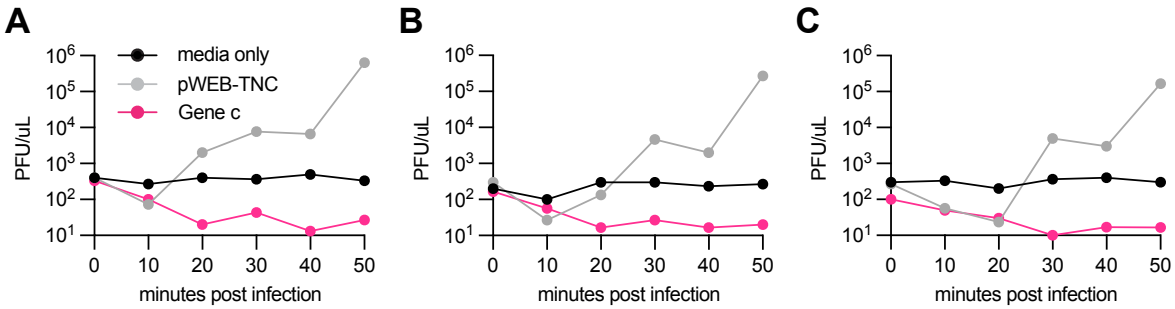


Figure 4.5 Gene c does not affect phage T4 adsorption

(**A-C**) Enumeration of T4 plaque forming units (PFU) in supernatants of infected cultures at different times after phage addition. Cultures of *E. coli* EC100 carrying pWEB-TNC or pFragmentD3-4 (expressing Gene c) were infected at MOI 0.01. Addition of T4 phage to media lacking bacteria was used as a control. Graphs for *n* = 3 independent experiments are shown in (**A-C**).

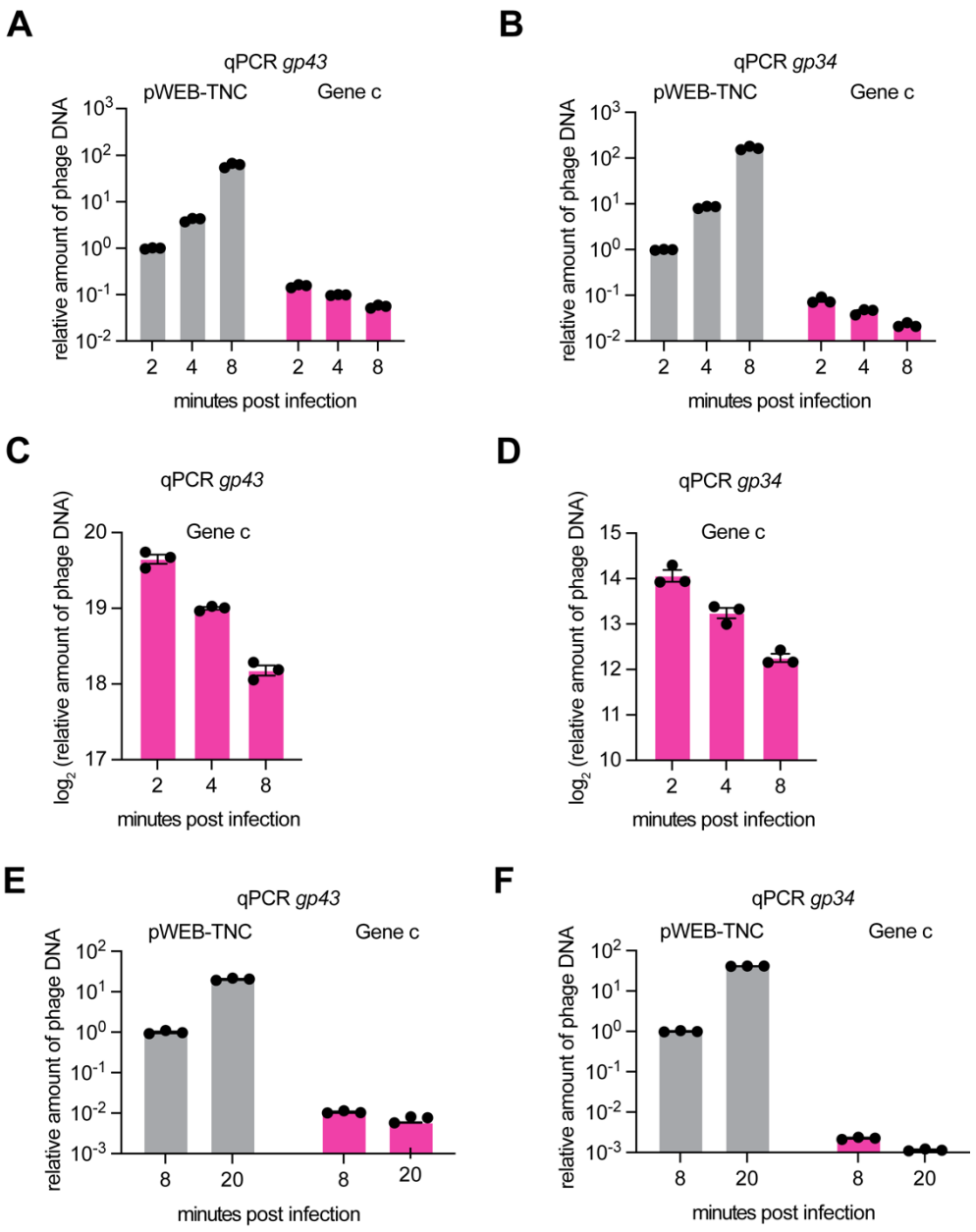


Figure 4.6 Quantitative PCR shows that Gene c inhibits T4 phage replication

(**A-B**) Quantitative PCR analysis of T4 DNA through amplification of the *gp43* and *gp34* genes, respectively. Viral DNA was extracted from infected *E. coli* EC100 cells carrying pWEB-TNC or pFragmentD3-4 (expressing Gene C) at 2, 4 and 8 minutes after the addition of phage at an MOI of 1. Fold-change values were calculated relative to the pWEB-TNC 2-minute time point. (**C-D**) Quantitative PCR results from (**A-B**), respectively, plotted using a $\log_2$ scale. (**E-F**) Same as (**A-B**) but using DNA extracted at 8- and 20-minutes post infection. Fold-change values were calculated relative to the pWEB-TNC 8-minute post-infection time point. For all quantitative PCR plots, mean ± SEM values are reported for $n = 3$ independent experiments.

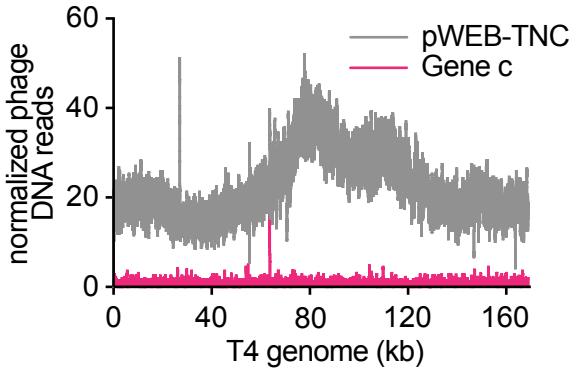


Figure 4.7 Next-generation sequencing of T4-infected cells shows that Gene c inhibits T4 phage replication

Normalized next-generation sequencing reads of T4 DNA, mapped to the viral genome. DNA for sequencing was obtained 8 minutes after infection of *E. coli* EC100 carrying pWEB-TNC or pFragmentD3-4 (expressing Gene C) at an MOI of 5.

4.4 Brig1 targets glucosylated DNA bases in bacteriophage T4

To understand how Brig1 affects T4 replication, we investigated T4 mutations that enable viral evasion of immunity. We isolated an “escaper” plaque that formed in the presence of Brig1 and confirmed that viruses from this plaque completely bypassed Brig1 defense (**Fig. 4.8-A**) and displayed a very similar pattern of DNA reads during infection in the presence or absence of Brig1 expression in *E. coli* hosts (**Fig. 4.8-B**). NGS of genomic DNA from this escaper phage revealed a single base-pair deletion within the T4 alpha-glucosyltransferase (*a-gt*) gene, resulting in a frameshift (escaper1, **Table 4.3**). Sequencing of PCR products obtained using DNA isolated from 18 additional phages that evaded Brig1 immunity showed additional mutations in *a-gt*, most of them frameshifts (**Table 4.3**).

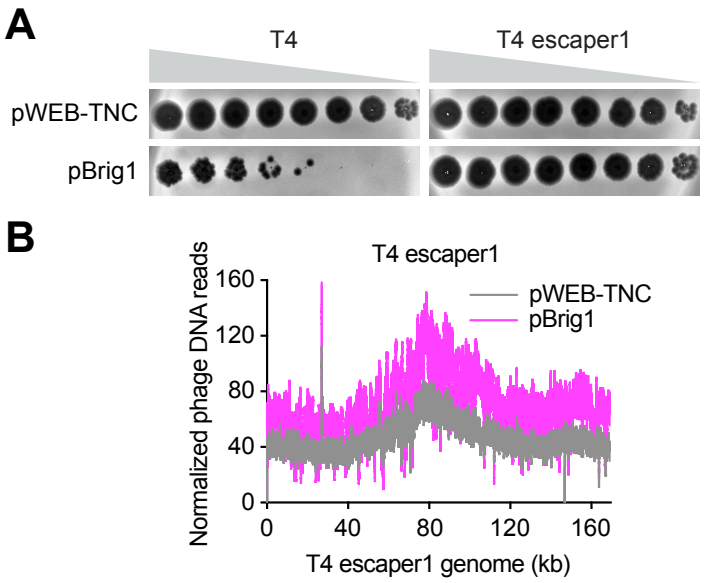


Figure 4.8 Isolation of a Brig1 escape phage that fully evades Brig1-mediated immunity

(**A**) Ten-fold serial dilutions of T4 or the Brig1 escape phage, T4 escaper1, on lawns of *E. coli* EC100 carrying pWEB-TNC or pBrig1. (**B**) Normalized next-generation sequencing reads of T4 escaper1 DNA, mapped to the viral genome. DNA for sequencing was obtained 8 minutes after infection of *E. coli* EC100 carrying pWEB-TNC or pBrig1 at an MOI of 5.

Table 4.3 Mutations in the *a-gt* gene of T4 phages that escape Brig1 immunity

escaper	Mutations in *a-gt* gene
T4-1	1 bp deletion of nt 713 (frameshift)
T4-2	1 bp insertion (A) after nt 294 (frameshift)
T4-3	nt 1031 G>A mutation (Gly344->Asp344)
T4-4	nt 793 C>G mutation (Pro265->Ala265) + 1 bp insertion (T) after nt 801 (frameshift)
T4-5	1 bp deletion of nt 294 (frameshift)
T4-6	nt 1072 G>T mutation (Glu358->TAA STOP)
T4-7	1 bp deletion of nt 294 (frameshift)
T4-8	1 bp insertion (T) after nt 996 (frameshift)
T4-9	nt 931 G>A mutation (Glu311->Lys311)
T4-10	nt 245 C>T mutation (Ser82->Phe82)
T4-11	1 bp insertion (A) after nt 294 (frameshift)
T4-12	1 bp insertion (A) after nt 294 (frameshift)
T4-13	1 bp insertion (A) after nt 294 (frameshift)
T4-14	1 bp deletion of nt 182 (frameshift)
T4-15	1 bp insertion (A) after nt 294 (frameshift)
T4-16	nt 926 A>G mutation (His309->Arg309)
T4-17	nt 977 G>A mutation (Gly326->Asp326)
T4-18	nt 419 A>G mutation (His140->Arg140)
T4-19	1 bp insertion (A) after nt 294 (frameshift)

We also generated an in-frame deletion of this gene, phage T4 Δ*a-gt*, which phenocopied the escaper1 mutation and led to viral propagation in the presence of Brig1 (**Fig. 4.9-A**). Finally, plasmid-borne expression of *a-gt* rescued *E. coli* from lysis by both T4 escaper1 and T4 Δ*a-gt* (**Fig. 4.9-A**), a result that demonstrates this gene is required for Brig1 defense. *a-gt* encodes alpha-glucosyltransferase (a-GT), which adds glucose in alpha-linkage to the 5-hydroxymethylcytosine (hmC) bases in the T4 genome (Lehman and Pratt, 1960; Sommer et al., 2004). During the T4 lytic cycle, all cytosines are replaced with hmC (Wyatt and Cohen, 1953) and subsequently glucosylated (**Fig. 4.9-B**). While a-GT adds a glucose residue in alpha-linkage to

~70% of the hmC bases, T4 also encodes a beta-glucosyltransferase (b-GT) that adds glucose in beta-linkage to the remaining ~30% (Lehman and Pratt, 1960; Sommer et al., 2004). The data obtained for T4 escaper1 and T4 Δ*a-gt*, phages that contain only beta-modified hmC residues, argues that beta-glucosyl-hmC residues are not required for Brig1 targeting. To corroborate this, we generated an in-frame deletion of *b-gt*, phage T4 Δ*b-gt*, and tested its susceptibility to Brig1. Plaque formation assays showed that the absence of beta-glucosyl-hmC nucleobases did not impede immunity (**Fig. 4.9-A**). In addition, overexpression of *b-gt*, which presumably increases the fraction of beta-glucosyl-hmC residues within the T4 genome, did not affect Brig1 immunity measured by plaque assays (**Fig. 4.9-A**). Altogether, these results demonstrate that Brig1 defense requires alpha-glucosylated hmC nucleobases in the infecting phage genome for efficient targeted defense.

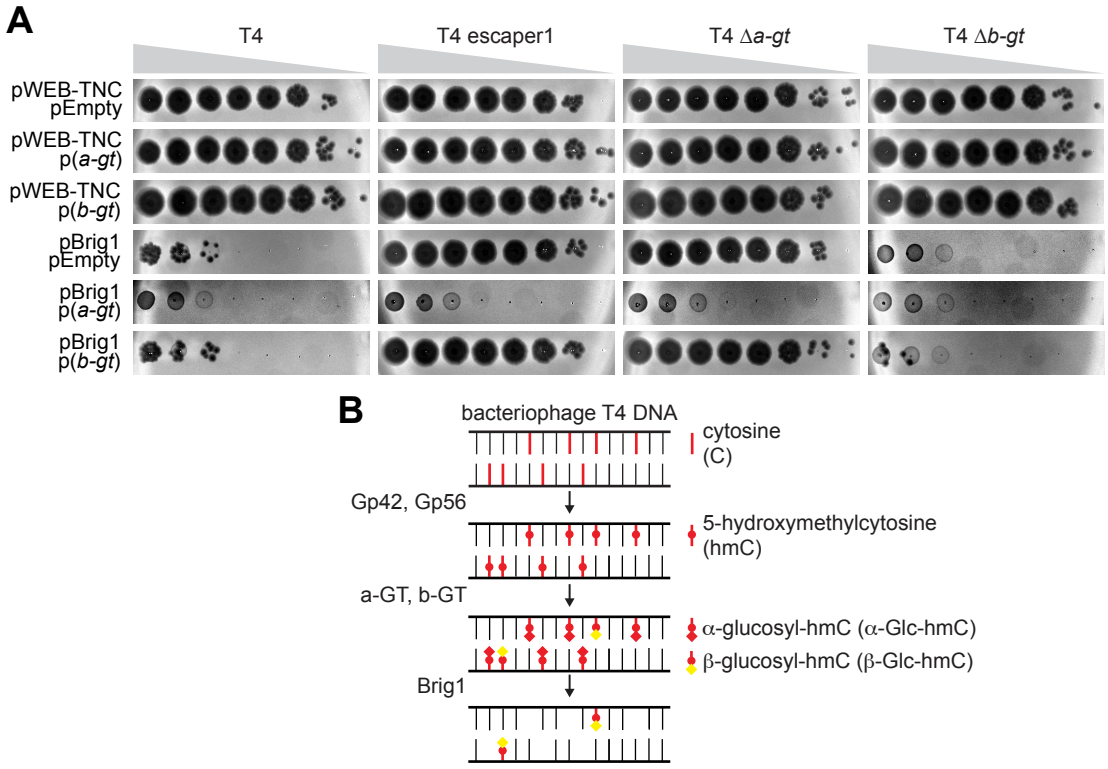


Figure 4.9 Brig1 targets alpha-glucosyl-hydroxymethylcytosine nucleobases in phage T4

(**A**) Ten-fold serial dilutions of different T4 phage stocks on lawns of *E. coli* EC100, each lawn carrying two plasmids: pWEB-TNC or pBrig1, and pEmpty, p(*a-gt*) or p(*b-gt*). Plaque images of one representative experiment from $n = 3$ independent experiments are shown. (**B**) Schematic of the cytosine modification pathway in phage T4, including a model for Brig1 immunity in which the DNA glycosylase recognizes and excises alpha-glucosyl-hydroxymethylcytosine nucleobases from the viral genome, generating abasic sites. T4 enzymes: Gp42, dCMP hydroxymethylase; Gp56, dCTPase; a-GT, alpha-glucosyltransferase; b-GT, beta-glucosyltransferase.

Since glycosylation of the T4 genome only occurs on hmC, we hypothesized that phage mutants that prevent synthesis of this DNA base would also be resistant to Brig1. To test this, we generated T4 phages lacking the hmC synthesis pathway, which employs two genes: (i) *gp42*, encoding a dCMP hydroxymethylase that adds a hydroxymethyl group to deoxycytidine monophosphate to generate hmC nucleotides (Chiu et al., 1976; Lamm et al., 1988), and (ii) *gp56*, encoding a dCTPase that depletes cytosines from the nucleotide pool (Carlson and Wiberg, 1983; Wiberg, 1967) (**Fig. 4.10-A**). In addition, phage T4 attacks unmodified cytosines present in the host DNA by both inhibiting the transcription of as well as degrading cytosine-containing DNA through the activities of the genes *alc* (Drivdahl and Kutter, 1990; Severinov et al., 1994) and *denB* (Carlson and Wiberg, 1983; Hirano et al., 2006), respectively (**Fig. 4.10-A**). We constructed a cytosine containing T4 mutant phage, T4(C), with the genotype Δ*alc* Δ*denB* Δ*gp56* Δ*gp42*, which has been shown to lack hmC in its genome (Bryson et al., 2015). This phage was resistant to Brig1 targeting, in contrast to the triple mutant Δ*alc* Δ*denB* Δ*gp56* phage that harbors both *gp42* to synthesize hmC nucleobases and *a-gt* to glycosylate the bases (**Fig. 4.10-B**).

Overexpression of *gp42*, but not *a-gt* alone, sensitized T4(C) to Brig1 targeting (**Fig. 4.10-B**). These results further corroborate that Brig1 targets alpha-glucosylated hmC nucleobases in the viral DNA to provide defense against T4.

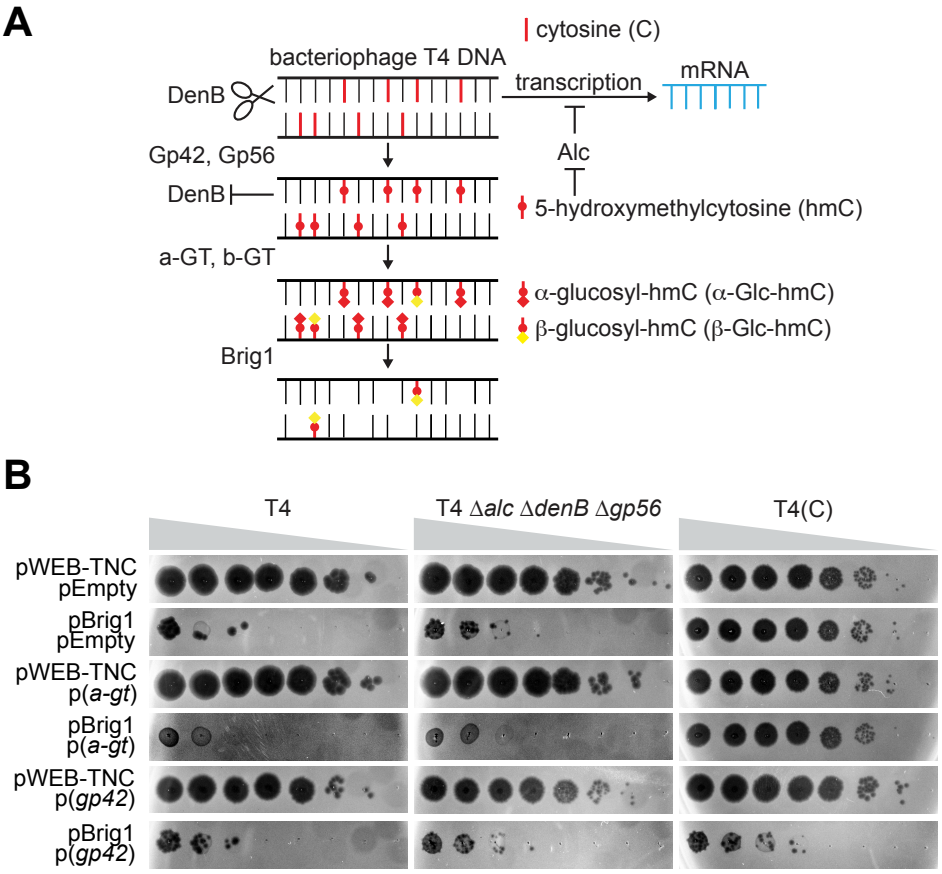


Figure 4.10 Brig1 targeting requires hydroxymethylcytosine nucleobases

(**A**) Schematic of the cytosine modification pathway in phage T4, showing the roles of T4 Alc and DenB, and including a model for Brig1 immunity in which the DNA glycosylase recognizes and excises alpha-glucosyl-hydroxymethylcytosine nucleobases from the viral genome, generating abasic sites. T4 enzymes: Alc, dC-specific premature transcriptional terminator; DenB, dC-specific ssDNA endonuclease; Gp42, dCMP hydroxymethylase; Gp56, dCTPase; a-GT, alpha-glucosyltransferase; b-GT, beta-glucosyltransferase. (**B**) Ten-fold serial dilutions of different T4 phage stocks on lawns of *E. coli* EC100, each lawn carrying two plasmids: pWEB-TNC or pBrig1, and pEmpty, p(*a-gt*) or p(*gp42*). Plaque images of one representative experiment from $n = 3$ independent experiments are shown.

4.5 Summary

In this chapter, I describe a methodology for screening environmental DNA libraries to uncover novel anti-phage defense systems in unsequenced bacteria. I hope this method will be broadly useful in driving discovery and generating new projects for current and future members of the Marraffini lab, as well as for the broader phage defense field. Using our approach, we discovered Brig1, a prokaryotic DNA glycosylase that provides immunity in *E. coli* against bacteriophage T4. By performing in vivo infection assays and by sequencing viral escape mutants, we established that Brig1 inhibits T4 viral replication by targeting alpha-glucosyl-hmC nucleobases present in T4 genomic DNA. Similar methods can also be implemented in the characterization of any future hits from functional screens. In the next chapter, I will delve deeper into a molecular and biochemical characterization of Brig1 and its mechanism of anti-phage immunity against T-even viruses.

CHAPTER 5. THE PROKARYOTIC DNA GLYCOSYLASE BRIG1 MEDIATES EXCISION OF HYPERMODIFIED VIRAL NUCLEOBASES

5.1 Background

In the previous chapter, I reported the discovery of Brig1, a prokaryotic DNA glycosylase that provides immunity against phage T4 by targeting alpha-glucosyl-hydroxymethylcytosine nucleobases, which are ubiquitous in the phage's DNA. "Hypermodification" of a phage's cytosine nucleobases has been shown to counter DNA targeting anti-phage defenses (Bryson et al., 2015; Hutinet et al., 2019; Liu et al., 2020b; Luria and Human, 1952; Revel, 1967; Vlot et al., 2018). Since DNA glycosylases are enzymes that typically function in base excision repair, Brig1 represents an example of a DNA repair protein being repurposed by bacteria for a novel immune purpose, with implications for bacteria-phage evolutionary conflicts. In this chapter, I provide a mechanistic and biochemical characterization of Brig1's immune activity against phage T4. Finally, in the next chapter, I will look at Brig1-mediated antiviral defense against diverse phages and at homologs of Brig1 that may possess roles in prokaryotic host-virus conflicts.

5.2 Brig1 is related to the uracil DNA glycosylase superfamily

To understand Brig1's enzymatic activity, we first performed an AlphaFold2 protein structure prediction (Mirdita et al., 2022) of Brig1, which generated a high

confidence structural model (**Fig. 5.1**). We used the Dali server to find structural homologs. The top hits were all uracil DNA glycosylases (Ahn et al., 2019; Barrett et al., 1998; Wibley et al., 2003), with the best match a family 4 uracil DNA glycosylase from the archaeon *Sulfolobus tokodaii* (Kawai et al., 2015) (Dali *Z* score 9.2) (Holm, 2022) (**Fig. 5.2**). Uracil DNA glycosylases recognize uracil bases in DNA (which may result from polymerase error or from cytosine deamination) and initiate base excision repair by hydrolyzing the glycosidic bond between the base and the deoxyribose sugar. Substrate recognition and catalysis are both facilitated by flipping the uracil base out of the double helix and into a pocket within the enzyme (Pearl, 2000; Schormann et al., 2014). Given the structural similarity, we hypothesized that Brig1 is a DNA glycosylase related to the uracil DNA glycosylase superfamily and that Brig1 excises alpha-glucosyl-hmC nucleobases, instead of uracil, from DNA backbones.

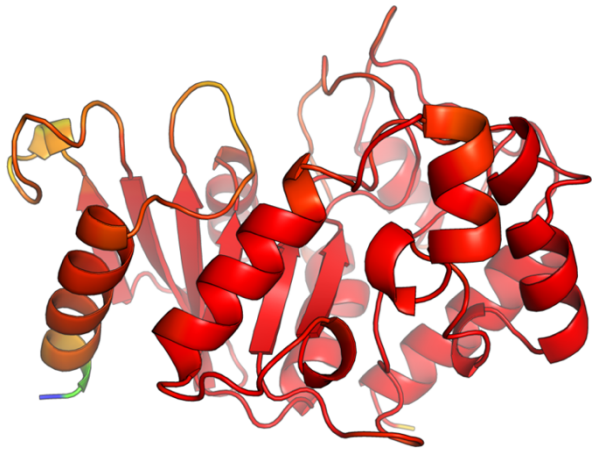


Figure 5.1 AlphaFold2 structure model of Brig1

AlphaFold2 structure of Brig1, colored by b-factors. Red to blue spectrum represents high to low confidence of secondary structure prediction.

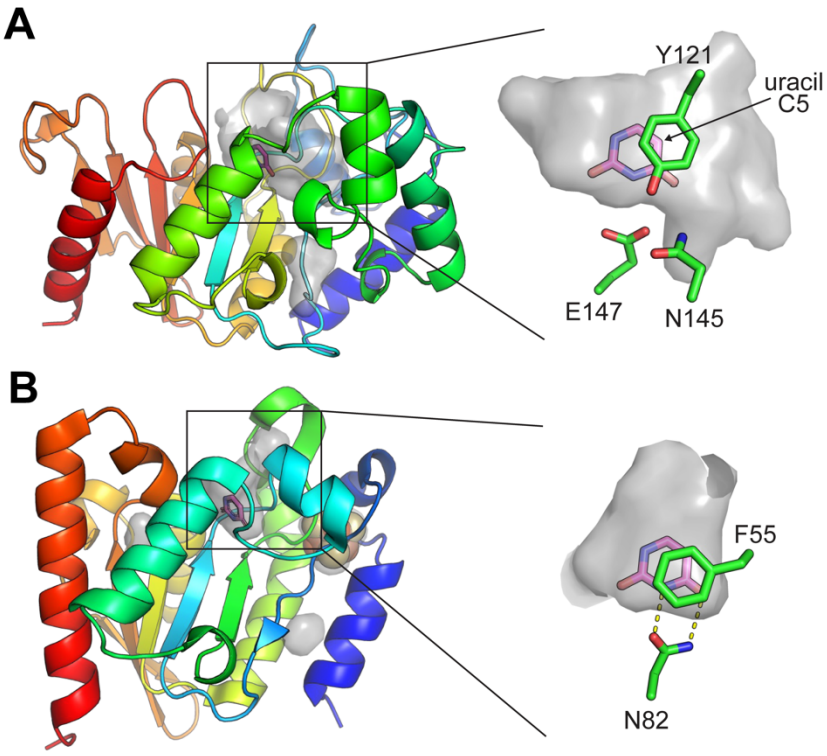


Figure 5.2 Brig1 structurally resembles uracil DNA glycosylases

(**A**) AlphaFold2 structure of Brig1, colored by N- to C-terminal (blue to red), with cavities shown in translucent grey. Inset; zoomed in view of putative glycosylase pocket, with uracil (pink-purple sticks) from PDB 4ZBY in (**B**) modeled into it, and showing the amino acid residues that could participate in substrate binding. (**B**) Crystal structure of a family 4 uracil DNA glycosylase from *Sulfolobus tokodaii* (PDB 4ZBY, a close structural homolog of Brig1) showing the uracil substrate (pink-purple sticks) in the glycosylase pocket and the Fe-S cluster (yellow and orange spheres). Structure is colored by N- to C-terminal (blue to red), with cavities shown in translucent grey. Inset, zoomed in view of the glycosylase pocket, showing the uracil substrate and the amino acid residues that participate in uracil binding.

5.3 Brig1 excises alpha-glucosyl-hydroxymethylcytosine nucleobases to generate abasic sites in ssDNA

Considering the homology with uracil DNA glycosylases and the finding that immunity is effective only when phage DNA contains alpha-glucosyl-hmC, we hypothesized that Brig1 removes these nucleobases, but not beta-glucosyl-hmC,

from the T4 genome. To test this prediction, we purified Brig1 and determined its activity on a 60-nt ssDNA oligonucleotide substrate containing a single hmC residue (**Fig. 5.3-A**), to which we enzymatically introduced the different glucosyl-hmC modifications present in T4 DNA (**Fig. 5.3-B**). To do this we used T4 a-GT and b-GT enzymes to add alpha- or beta-glucosyl groups, respectively, to a commercially synthesized oligonucleotide containing a single hmC residue within an MfeI restriction site (**Fig. 5.3-A**), either before or after annealing it to a complementary oligonucleotide to generate a dsDNA substrate. Glucosylation was confirmed by testing for the resistance of this dsDNA to MfeI digestion, a restriction endonuclease that can cleave hmC- but not glucosyl-hmC-containing sequences (Liu et al., 2020a) (**Fig. 5.3-C**). The modified ssDNA oligonucleotides were incubated with Brig1 to determine its DNA glycosylase activity using an aldehyde-reactive fluorescent probe that can detect abasic sites (**Fig. 5.4**). This probe detects the aldehyde-containing ring-opened form of the deoxyribose that is in equilibrium with the closed-ring form after removal of the base. As a positive control, we treated an equivalent uracil-containing oligonucleotide (**Fig. 5.3-DE**) with hSMUG1, a previously characterized human uracil DNA glycosylase (Masaoka et al., 2003). As with hSMUG1 treatment of the uracil-containing oligonucleotide, the primary product of Brig1 was a full-length oligonucleotide containing an abasic site, generated by excision of the alpha- but not the beta-glucosyl-hmC nucleobase (**Fig. 5.4**). Also, just as with hSMUG1, a small fraction of the target substrate was cleaved at the position of the modified base (**Fig. 5.4**), suggesting that both hSMUG1 and Brig1 may have a very weak

apurinic/apyrimidinic (AP) lyase activity on ssDNA substrates in addition to their DNA glycosylase activity (Alexeeva et al., 2019).

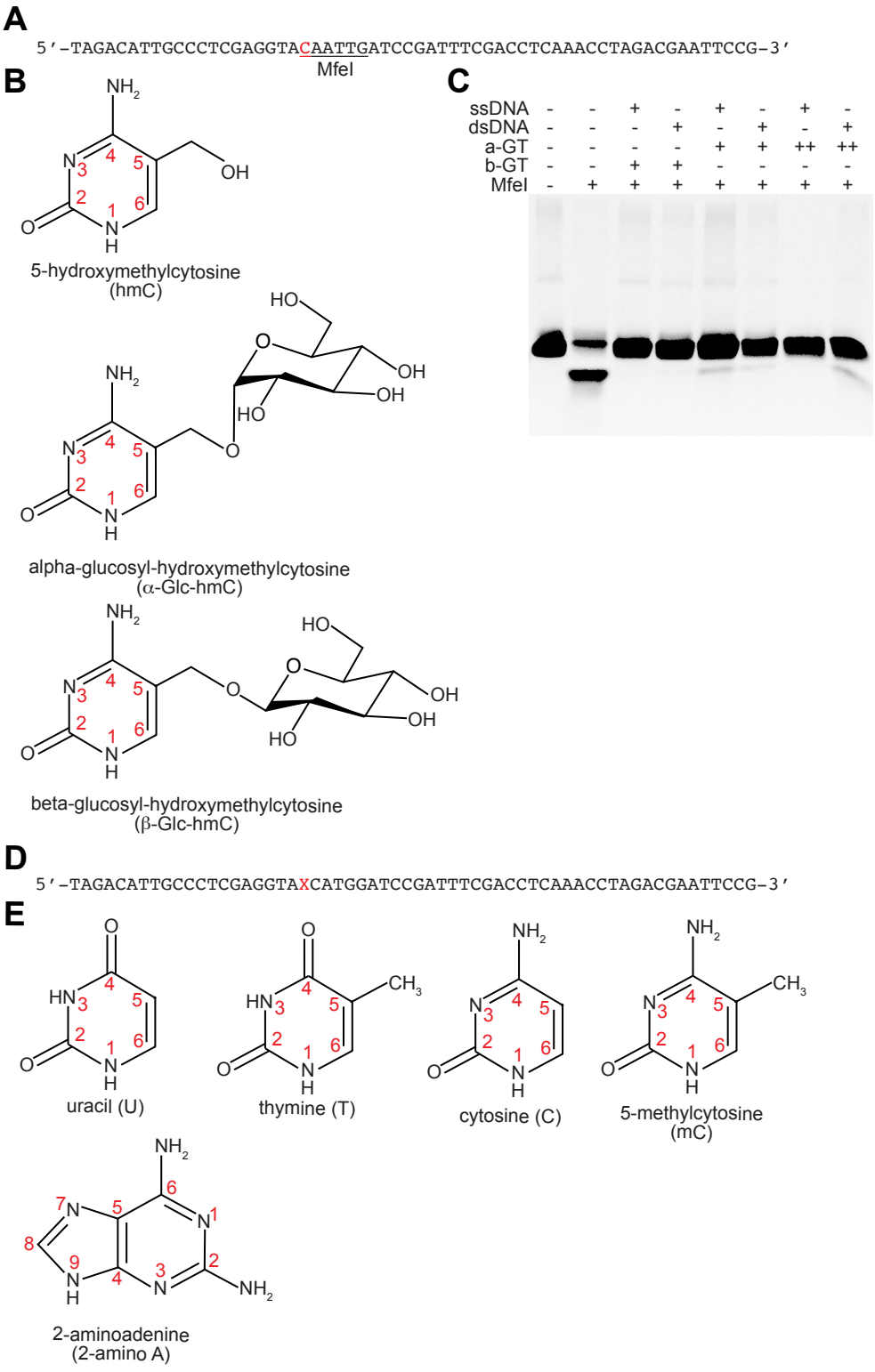


Figure 5.3 Oligonucleotides and nucleobases used in Brig1 DNA glycosylase assays

(**A**) Sequence of the 60-nt single-stranded oligonucleotide used for testing the DNA glycosylase activity of Brig1. The cytosine in red was synthesized as 5-hydroxymethylcytosine and subsequently alpha- or beta-glucosylated. The MfeI restriction site is underlined. MfeI digestion was used to confirm glucosylation after annealing a complementary bottom strand oligonucleotide to create a dsDNA substrate for the MfeI enzyme. (**B**) Chemical structures of 5-hydroxymethylcytosine as well as alpha- and beta-glucosyl-hydroxymethylcytosine nucleobases found in phage T4 DNA. (**C**) MfeI digestion of dsDNA substrates generated by annealing the oligonucleotide shown in (**A**) containing a 5-hydroxymethylcytosine nucleobase with a complementary bottom strand oligonucleotide. In each case, prior to MfeI digestion of the dsDNA, either the pre-annealed ssDNA oligonucleotide or the annealed dsDNA substrate was either untreated or treated with a low (+) or high (++) concentration of T4 alpha- or T4 beta-glucosyltransferase (a-GT or b-GT, respectively). (**D**) Sequence of the 60-nt single-stranded oligonucleotide used for testing the DNA glycosylase activity of Brig1 or hSMUG1. The red X was replaced by the nucleobases shown in (**E**). (**E**) Chemical structures of different nucleobases used to test the specificity of Brig1.

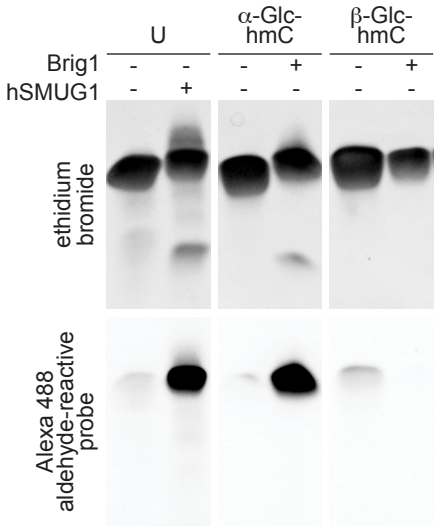


Figure 5.4 Brig1 excises alpha-glucosyl-hydroxymethylcytosine nucleobases in ssDNA to generate abasic sites

Polyacrylamide gel electrophoresis of 60-nt single-stranded oligonucleotides containing a single modified base, incubated with either hSMUG1 or Brig1 at 37°C overnight, and then treated with an aldehyde-reactive Alexa 488 fluorescent probe to label and detect abasic sites. The same gel was stained with ethidium bromide to detect ssDNA. U, uracil; α-Glc-hmC, alpha-glucosyl-hydroxymethylcytosine; β-Glc-hmC, beta-glucosyl-hydroxymethylcytosine.

We next tested the substrate specificity of Brig1 using a panel of commercially synthesized single-stranded oligonucleotides, each containing a different modified nucleotide at the same position (**Fig. 5.3-DE**). After overnight incubation with Brig1, substrates were run on a polyacrylamide gel (**Fig. 5.5-A**) or heated with NaOH prior to gel electrophoresis, which accelerates the cleavage of the DNA backbone at abasic sites via beta-elimination (Alexeeva et al., 2019; Bailly and Verly, 1987), allowing clear visualization of enzymatic base excision activity (**Fig. 5.5-B**). Brig1 exhibited robust base excision activity on the alpha-glucosyl-hmC containing substrate but not on ssDNA substrates harboring beta-glucosyl-hmC, hmC, 5-methylcytosine, or 2-aminoadenine (**Fig. 5.5-B**). We further confirmed that Brig1 generates an abasic site in its target oligonucleotide by incubating the Brig1-treated substrate with a commercially available AP endonuclease, NEB Endonuclease IV, an enzyme which cleaves the DNA backbone adjacent to an abasic site, just as with heat- and base-promoted beta-elimination (Bailly and Verly, 1987; Thompson and Cortez, 2020). Treatment with Endonuclease IV produced the same ssDNA cleavage pattern as when the Brig1-treated oligonucleotide was heated with NaOH to promote beta-elimination (**Fig. 5.5-C**), confirming that Brig1 activity generates an abasic site in the target oligonucleotide.

Interestingly, we observed weak base excision activity after treating a uracil-containing ssDNA with Brig1 (**Fig. 5.5-B**), confirming that Brig1 is related to the uracil DNA glycosylase superfamily and possesses a very weak uracil excision activity that

is secondary to its robust activity on alpha-glucosyl-hmC. Together, our results demonstrate that Brig1 is a DNA glycosylase that excises alpha-glucosyl-hmC nucleobases from ssDNA to generate abasic sites, with a high level of stereoisomeric specificity. Its weak secondary activity on uracil suggests Brig1 has evolved divergently from canonical members of the uracil DNA glycosylase superfamily.

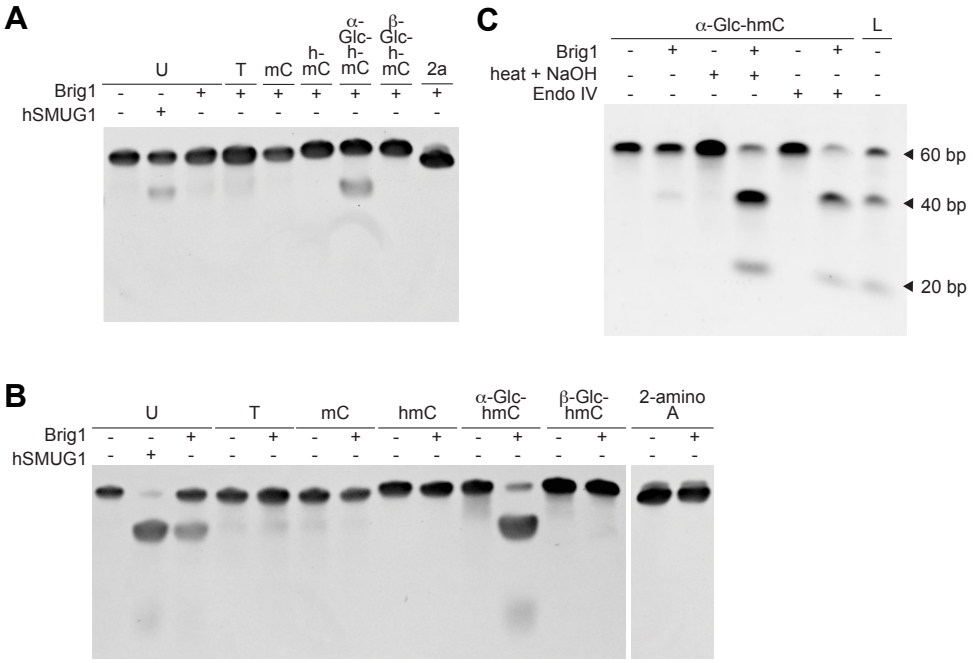


Figure 5.5 Brig1 has base excision specificity for alpha-glucosyl-hydroxymethylcytosine with a weak secondary activity on uracil

(**A-B**) Polyacrylamide gel electrophoresis of 60-nt single-stranded oligonucleotides containing a single modified base, incubated with either hSMUG1 or Brig1 at 37°C overnight, and then subjected to gel electrophoresis (**A**) or treated with NaOH and heat for 30 minutes prior to gel electrophoresis (**B**). Gels were stained with ethidium bromide to detect ssDNA. U, uracil; T, thymine, mC, 5-methylcytosine; hmC, 5-hydroxymethylcytosine; α-Glc-hmC, alpha-glucosyl-hydroxymethylcytosine; β-Glc-hmC, beta-glucosyl-hydroxymethylcytosine; 2-aminoA or 2A, 2-aminoadenine. **(C)** Polyacrylamide gel electrophoresis of the 60-nt single-stranded oligonucleotide containing an alpha-glucosyl-hydroxymethylcytosine nucleobase (α-Glc-hmC), incubated with or without Brig1 at 37°C overnight and then incubated with and without 50 units of NEB Endonuclease IV at 37°C for 4 hours or heated with NaOH for 30 minutes prior to gel electrophoresis. Gels were stained with ethidium bromide to detect ssDNA. L, ssDNA size ladder.

5.4 Mass spectrometry confirms that Brig1 activity generates an abasic site

Next, we used high resolution mass spectrometry (MS) to confirm the removal of the target nucleobase and the formation of an abasic site in ssDNA through the DNA glycosylase activity of Brig1. We treated an 18-nt ssDNA oligonucleotide containing either a single uracil, hmC or alpha-glucosyl-hmC nucleobase (**Fig. 5.6-AB**) to overnight incubation with hSMUG1 or Brig1. MS detects an abasic site as a decrease in mass of the original 18-nt oligonucleotide, equivalent to the loss of the excised target nucleobase and the gain of a water molecule (**Fig. 5.6-C**). The exact mass is often accompanied by multiple heavier peaks, representing adducts introduced by the method of mass measurement (**Fig. 5.6-DEFGHIJKL**). With hSMUG1 and Brig1 treatment of the uracil- and alpha-glucosyl-hmC containing oligonucleotides, respectively, we recorded, in each case, a strong primary MS peak with a mass that corresponds to the loss of the nucleobase and the gain of water, i.e., an abasic site. These experiments confirm that Brig1 activity on target nucleobases generates abasic sites in the DNA.

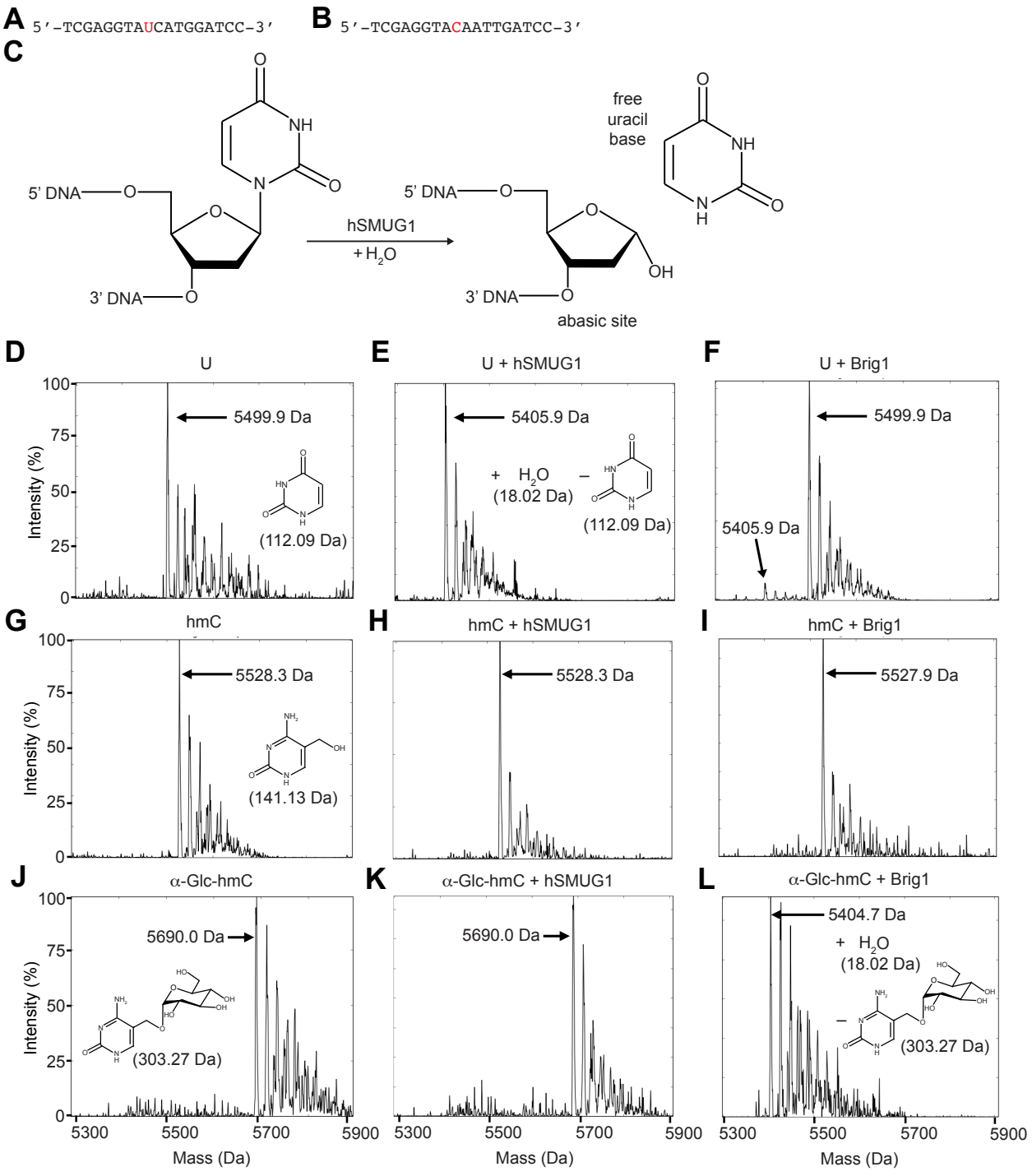

A 5′-TCGAGGTAUCATGGATCC-3′
B 5′-TCGAGGTACAATTGATCC-3′
C
5′ DNA
3′ DNA
hSMUG1
+ H2O
free uracil base
abasic site
D
U
5499.9 Da
(112.09 Da)
E
U + hSMUG1
5405.9 Da
+ H2O (18.02 Da) –
(112.09 Da)
F
U + Brig1
5499.9 Da
5405.9 Da
G
hmC
5528.3 Da
(141.13 Da)
H
hmC + hSMUG1
5528.3 Da
I
hmC + Brig1
5527.9 Da
J
α-Glc-hmC
5690.0 Da
(303.27 Da)
K
α-Glc-hmC + hSMUG1
5690.0 Da
L
α-Glc-hmC + Brig1
5404.7 Da
+ H2O (18.02 Da)
– (303.27 Da)
Intensity (%)
Mass (Da)

Figure 5.6 Mass spectrometry of a Brig1-generated abasic site

(**A**) Sequence of the uracil-containing, 18-nt single-stranded oligonucleotide used for mass spectrometry (MW 5500.6 Da). (**B**) Sequence of the 5-hydroxymethylcytosine (hmC)-containing, 18-nt single-stranded oligonucleotide used for mass spectrometry. The cytosine in red was synthesized as hmC (MW 5528.7 Da) and subsequently alpha-glucosylated (MW 5690.9 Da) by purified T4 alpha-glucosyltransferase. (**C**) Generation of an abasic site by a DNA glycosylase through excision and release of its target nucleobase from the DNA backbone. Adapted from Thompson and Cortez (2020) and reproduced with permission from Elsevier. (**D-L**) Zero-charge mass spectra from high resolution mass spectrometry of oligonucleotides from (**A**) or (**B**) incubated with no enzyme, hSMUG1 or Brig1 at 37°C overnight. Major mass peaks are indicated. Relevant nucleobases are shown in the appropriate panels.

5.5 Mutations in the catalytic pocket of Brig1 abrogate base excision activity

Consistent with the different substrate specificities, the nucleotide binding pocket is predicted to be much larger in the Brig1 model (**Fig. 5.2-A**) than in the *S. tokodaii* uracil DNA glycosylase (**Fig. 5.2-B**), with extra space adjacent to the C5 position of the pyrimidine where the additional alpha-glucosyl-hydroxymethyl group would protrude (**Fig. 5.2-A** and **Fig. 5.3-B**). To test if this putative binding pocket is important for Brig1 activity, we mutated amino acids predicted to outline this area: Y121, E147 and N145 (**Fig. 5.2-A**). Based on the structure of other related glycosylases, Y121 would stack against the flipped-out base (as is the case for F55 in *S. tokodaii* uracil DNA glycosylase, **Fig. 5.2-B**), while E147 would form hydrogen bonds to its Watson:Crick face. Because this residue is often asparagine rather than glutamate (Aravind and Koonin, 2000) (for example N82 in *S. tokodaii* uracil DNA

glycosylase, **Fig. 5.2-B**), we also considered the N145 residue (**Fig. 5.2-A**). We substituted each of these residues for alanine and tested the Brig1 mutants for their ability to provide immunity against T4. We found that while the N145A substitution did not affect Brig1-mediated immunity, the Y121A and E147A mutants failed to reduce T4 plaque formation (**Fig. 5.7-A**). Substitution of E147 for glutamine, which eliminates the negative charge on this residue, also abrogated Brig1 activity. Finally, we tested a purified Brig1 double mutant (Y121A, E147A) for base excision activity in vitro and found that the binding pocket mutations abrogated Brig1's enzymatic activity (**Fig. 5.7-BC**). These results demonstrate that the putative DNA glycosylase catalytic pocket of Brig1 is essential for base excision activity and defense against phage T4.

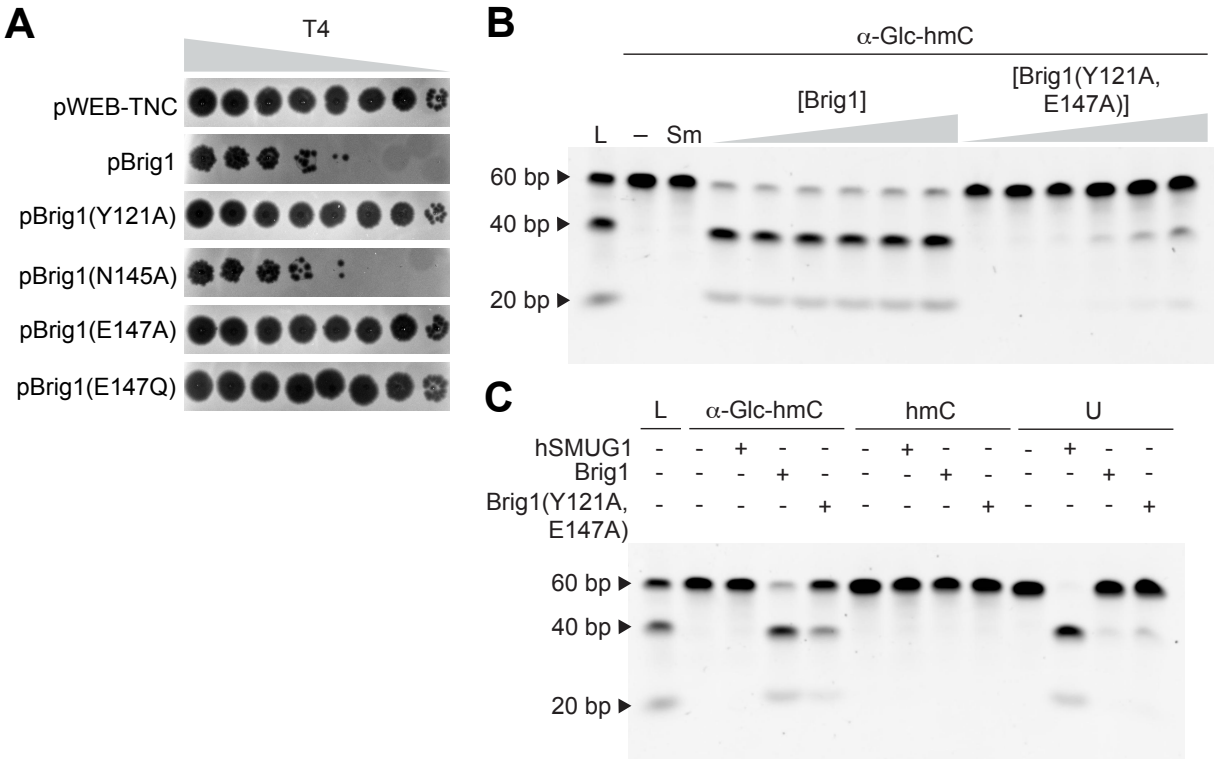


Figure 5.7 Mutations in the DNA glycosylase pocket of Brig1 abrogate enzymatic activity

(**A**) Ten-fold serial dilutions of phage T4 on lawns of *E. coli* EC100 carrying pWEB-TNC, pBrig1 or pBrig1 harboring substitutions in the amino acids thought to participate in substrate binding. Plaque images of one representative experiment from $n = 3$ independent experiments are shown. (**B**) Polyacrylamide gel electrophoresis (PAGE) of the 60-nt single-stranded oligonucleotide from Fig. 5.3-A containing an alpha-glucosyl-hydroxymethylcytosine nucleobase (α-Glc-hmC), incubated with Brig1 or a Brig1(Y121A, E147A) double mutant at 37°C for 30 minutes and then heated with NaOH for another 30 minutes prior to PAGE. Gels were stained with ethidium bromide to detect ssDNA. Enzyme concentrations: 50, 100, 200, 400, 800 and 1600 nM; Sm, 5 units of NEB hSMUG1; –, no enzyme; L, ssDNA size ladder. (**C**) PAGE of the 60-nt single-stranded oligonucleotide from Fig. 5.3-A containing an alpha-glucosyl-hydroxymethylcytosine nucleobase (α-Glc-hmC), the oligonucleotide containing 5-hydroxymethylcytosine (hmC) in place of α-Glc-hmC, and the oligonucleotide from Fig. 5.3-D containing a uracil (U) nucleobase. Oligonucleotides were incubated with 5 units of NEB hSMUG1, 1 µM Brig1 or 1 µM Brig1(Y121A, E147A) at 37°C for 2 hours and then heated with NaOH for another 30 minutes prior to PAGE. L, ssDNA size ladder.

5.6 Brig1 generates abasic sites in dsDNA

We next sought to test Brig1 activity on dsDNA substrates, especially since Brig1 provides immunity against phage T4, which is a dsDNA virus. We generated dsDNA oligonucleotides, by annealing our previously used ssDNA oligonucleotide containing a single hmC or alpha-glucosyl-hmC nucleobase to a complementary bottom strand ssDNA oligonucleotide without any modified nucleobases (**Fig. 5.8-AC**). We generated an equivalent dsDNA substrate with uracil in the top strand (**Fig. 5.8-B**), to serve as a positive control for base excision activity assayed using the uracil DNA glycosylase hSMUG1. Incubation with hSMUG1 of the dsDNA substrate

containing uracil showed a slight upward shift of the DNA band during polyacrylamide gel electrophoresis (PAGE) only when the hSMUG1-treated dsDNA was heated with NaOH prior to electrophoresis (**Fig. 5.8-D**). Since nicked linear duplex DNA runs slower on a gel than intact linear dsDNA (Kuhn et al., 2002), this result is indicative of a DNA backbone nick being generated via beta-elimination at the abasic site created by hSMUG1. Strand cleavage by beta-elimination at the abasic site was confirmed by Urea-PAGE, which resolves dsDNA substrates into their constituent nicked and intact ssDNA strands (**Fig. 5.8-E**). We observed both the slight upward gel shift with Brig1 on the alpha-glucosyl-hmC-containing dsDNA during PAGE and the accompanying ssDNA cleavage pattern during Urea-PAGE, only when the Brig1-treated dsDNA was heated with NaOH prior to gel loading (**Fig. 5.8-DF**). Brig1's base excision activity thus extends to both ssDNA and dsDNA substrates. No Brig1 activity was observed on the hmC-containing dsDNA (**Fig. 5.8-G**) and a very weak activity was detected on the uracil-containing dsDNA (**Fig. 5.8-E**). Importantly, we did not observe any gel shift or ssDNA cleavage for either hSMUG1 or Brig1 on their target dsDNAs without heat and NaOH (**Fig. 5.8-DEF**), indicating a lack of AP lyase activity under these assay conditions.

To gain a more comprehensive insight into Brig1 activity on dsDNA, we probed whether Brig1-mediated base excision of nearby alpha-glucosyl-hmC nucleobases on opposite strands of a dsDNA duplex (**Fig. 5.8-A**) may trigger spontaneous or enzyme-directed dsDNA cleavage akin to a restriction endonuclease. The target

dsDNA was generated by annealing two complementary ssDNA strands, each containing a single hmC residue, which were glucosylated with T4 a-GT prior to annealing (**Fig. 5.8-AC**). A control experiment was performed with a dsDNA containing two uracil residues a few nucleobases apart on opposite strands (**Fig. 5.8-B**). Incubation of the uracil-containing dsDNA with hSMUG1, followed by heat treatment with NaOH, exhibited a cleavage pattern during PAGE that was consistent with beta-elimination-mediated strand cleavage at the abasic sites generated by hSMUG1 (**Fig. 5.8-BH**). Importantly, Brig1 produced a pronounced cleavage pattern when the dsDNA with two alpha-glucosyl-hmC residues on opposite strands was heated with NaOH post enzymatic treatment (**Fig. 5.8-AI**). Expectedly, no cleavage was observed with the equivalent, non-glucosylated hmC-containing dsDNA oligonucleotide (**Fig. 5.8-J**). Furthermore, dsDNA cleavage was not observed with Brig1 (or hSMUG1) in the absence of heat and NaOH. Brig1 is therefore primarily a monofunctional DNA glycosylase that generates abasic sites, with little to no strand cleavage activity on dsDNA, driving an alternative antiviral mechanism of action that attacks phage DNA differently to DNA-cleaving restriction endonucleases, such as MfeI (**Fig. 5.8-C**), that create double-strand DNA breaks.

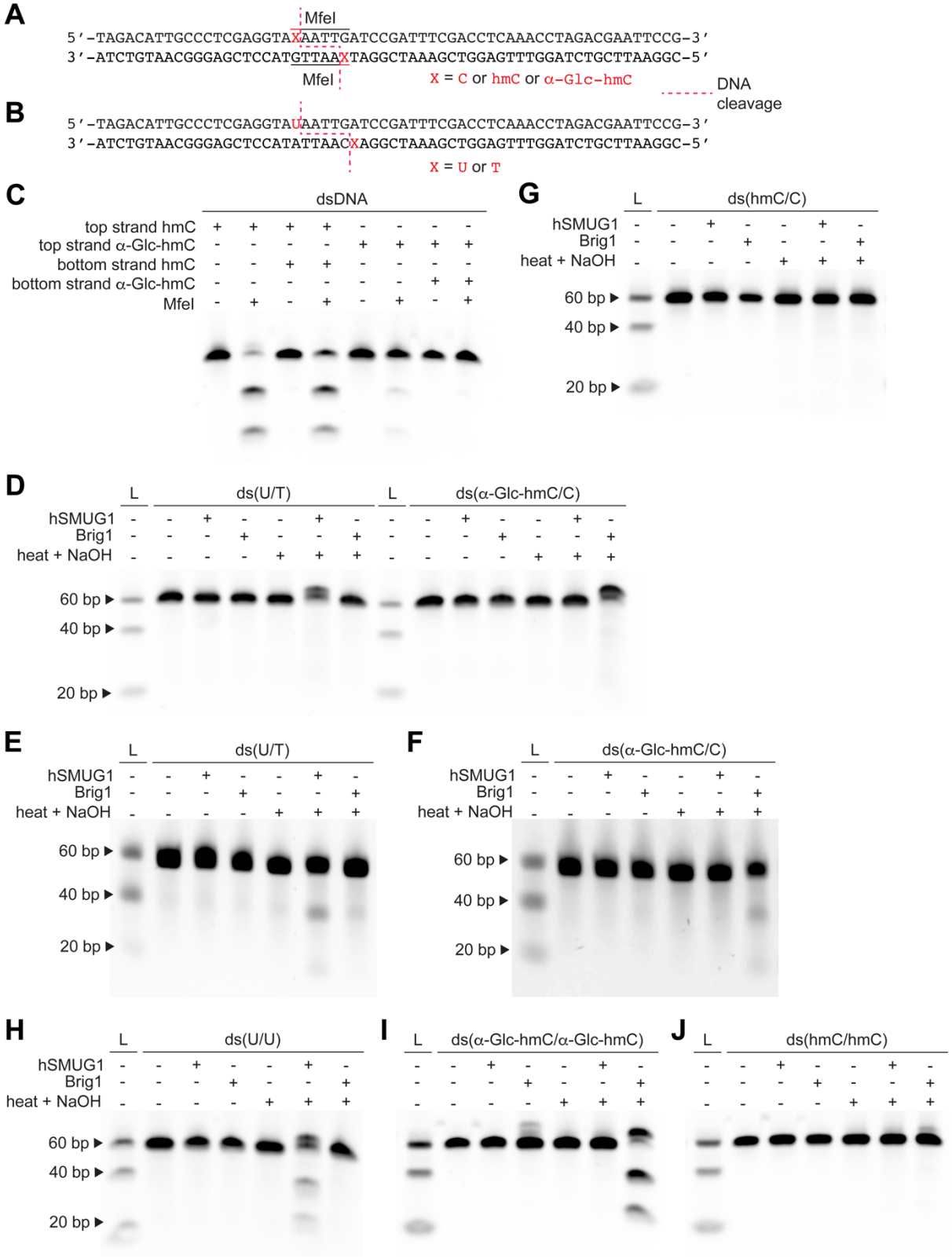

A
MfeI
5′-TAGACATTGCCCTCGAGGTAXAATTGATCCGATTTCGACCTCAAACCTAGACGAATTCCG-3′
3′-ATCTGTAACGGGAGCTCCATGTTAAXTAGGCTAAAGCTGGAGTTTGGATCTGCTTAAGGC-5′
MfeI
X = C or hmC or α-Glc-hmC
DNA cleavage
B
5′-TAGACATTGCCCTCGAGGTAUAATTGATCCGATTTCGACCTCAAACCTAGACGAATTCCG-3′
3′-ATCTGTAACGGGAGCTCCATATTAACXAGGCTAAAGCTGGAGTTTGGATCTGCTTAAGGC-5′
X = U or T
C
dsDNA
top strand hmC
top strand α-Glc-hmC
bottom strand hmC
bottom strand α-Glc-hmC
MfeI
G
L
ds(hmC/C)
hSMUG1
Brig1
heat + NaOH
60 bp
40 bp
20 bp
D
L
ds(U/T)
ds(α-Glc-hmC/C)
E
F
H
ds(U/U)
I
ds(α-Glc-hmC/α-Glc-hmC)
J
ds(hmC/hmC)

Figure 5.8 Brig1 generates abasic sites in dsDNA

(**A**) Sequence of the 60-nt double-stranded oligonucleotide used for testing the DNA glycosylase activity of Brig1. The base X in red on the top strand was synthesized as 5-hydroxymethylcytosine (hmC) and subsequently alpha-glucosylated (α-Glc-hmC); X in red on the bottom strand was synthesized as cytosine (C) or as hmC that was subsequently alpha-glucosylated (α-Glc-hmC). The MfeI restriction site is underlined. MfeI digestion was used to confirm glucosylation. Dotted lines indicate sites of DNA strand cleavage via beta-elimination at an abasic site. (**B**) Sequence of the 60-nt double-stranded oligonucleotide used for testing the uracil DNA glycosylase activity of hSMUG1 and Brig1. The base X in red on the top strand was synthesized as uracil (U); X in red on the bottom strand was thymine (T) or U. Dotted lines indicate sites of DNA strand cleavage via beta-elimination at an abasic site. (**C**) MfeI digestion (NEB, 40 units) of the dsDNA oligonucleotide (500 ng) from (**A**) containing hmC or α-Glc-hmC in the top and/or bottom strands. dsDNA substrates were generated by annealing untreated or alpha-glucosyltransferase-treated top and bottom strand ssDNA oligonucleotides. (**D**) Polyacrylamide gel electrophoresis (PAGE) of dsDNA oligonucleotides from (**A**) and (**B**) incubated with either hSMUG1 or Brig1 at 37°C overnight, and then subjected to gel electrophoresis or heated with NaOH for 30 minutes prior to electrophoresis. Gels were stained with ethidium bromide. ds(U/T), dsDNA oligonucleotide from (**B**) containing U and T on top and bottom strands, respectively; ds(α-Glc-hmC/C), dsDNA oligonucleotide from (**A**) containing α-Glc-hmC and C on top and bottom strands, respectively; L, dsDNA size ladder. (**E-F**) Urea-PAGE gels of the reactions in (**D**); L, ssDNA size ladder. (**G-J**) Same as (**D**) but with different oligonucleotides: (**G**) ds(hmC/C), dsDNA oligonucleotide from (**A**) containing hmC and C on top and bottom strands, respectively; (**H**) ds(U/U), dsDNA oligonucleotide from (**B**) containing U on both top and bottom strands; (**I**) ds(α-Glc-hmC/α-Glc-hmC), dsDNA oligonucleotide from (**A**) containing α-Glc-hmC on both top and bottom strands; (**J**) ds(hmC/hmC), dsDNA oligonucleotide from (**A**) containing hmC on both top and bottom strands. L, dsDNA size ladder. For all assays, 1 μM oligonucleotide, 1 μM of Brig1 and 5 units of NEB hSMUG1 were used unless otherwise stated.

5.7 Brig1 degrades T4 phage DNA in vitro

To investigate the molecular mechanism by which Brig1 restricts T4 infection, we tested the effect of purified Brig1 on T4 phage DNA in vitro. We incubated wild-type and escaper1 viral DNA with Brig1 for 30 minutes at 37°C and visualized the

products via agarose gel electrophoresis. While increasing concentrations of Brig1 caused a mobility shift in escaper1 DNA, it resulted in partial degradation of the wild-type DNA (**Fig. 5.9-A**). We attribute this observed DNA degradation to uncatalyzed, spontaneous beta-elimination rather than a significant enzymatic AP lyase activity, given our previous results and the reactive nature of abasic sites. Similar degradation of uracil-rich DNA was observed following a 30-minute incubation with the uracil DNA glycosylase hSMUG1, only when the substrate contained a very high concentration of deoxyuridines (**Fig. 5.10**). For Brig1 incubation, heating of the reaction products to 65°C before electrophoresis eliminated the mobility shift of the T4 escaper1 DNA (most likely due to denaturation of Brig1 and release of Brig1-bound non-target DNA) and resulted in complete degradation of the wild-type phage DNA due to heat-promoted beta-elimination at the abasic sites generated by Brig1 (**Fig. 5.9-B**). Treatment with SDS prior to gel electrophoresis of the DNA (to denature the Brig1 enzyme) also eliminated the mobility shift but did not cause further DNA degradation (**Fig. 5.9-C**). Importantly, mutations in the DNA glycosylase pocket of Brig1 (**Fig. 5.2-A**) abrogated T4 DNA degradation in vitro, both with and without heat (**Fig. 5.11**). In all experiments, we used pWEB-TNC cosmid DNA as a negative control, which showed similar mobility shifts and resistance to degradation as the T4 escaper1 DNA. We believe that these mobility shifts are a result of Brig1 binding to unmodified DNA (T4 escaper1 and pWEB-TNC cosmid).

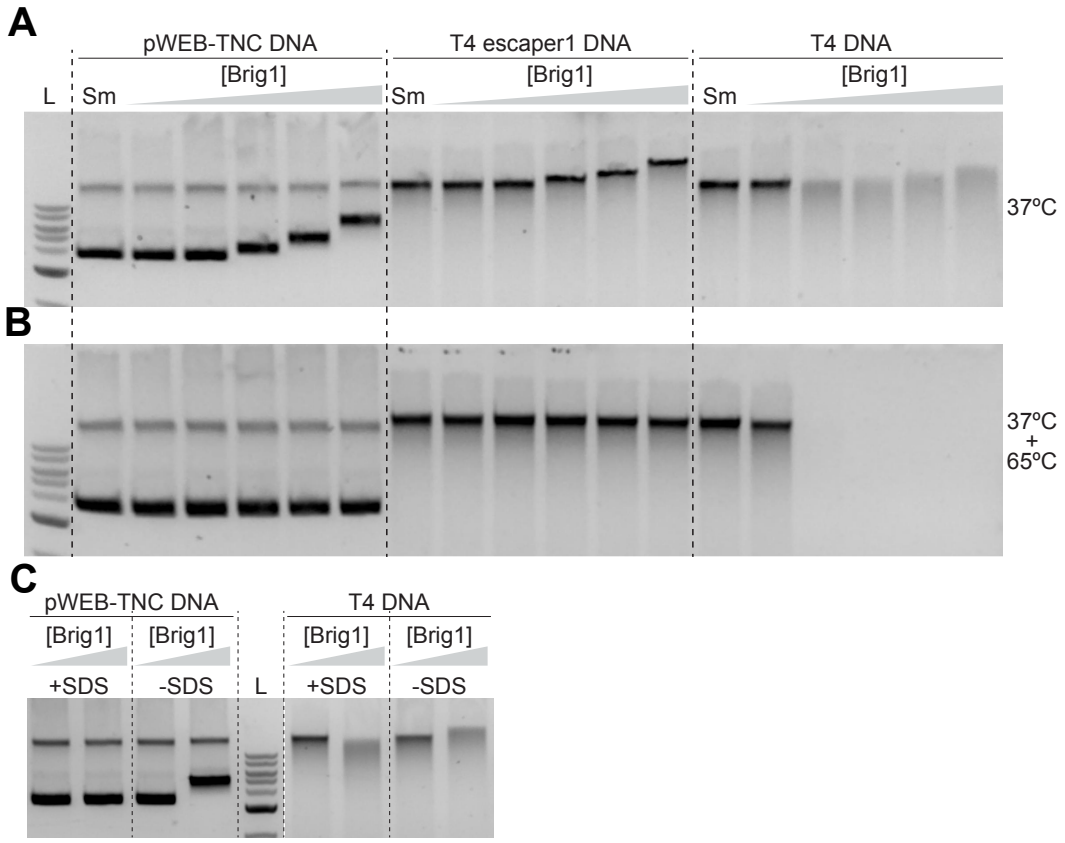


Figure 5.9 Brig1 degrades T4 phage DNA in vitro

(**A**) Agarose gel electrophoresis of T4, T4 escaper1 or pWEB-TNC DNA (500 ng) treated with increasing concentrations (2, 20, 200, 400, 800 nM) of Brig1 or with 10 units of NEB hSMUG1 (Sm) for 30 minutes at 37°C. L, DNA size ladder. (**B**) Same as (**A**) but heated for an additional 20 minutes at 65°C prior to gel electrophoresis. (**C**) Agarose gel electrophoresis of T4 or pWEB-TNC DNA (500 ng) treated with increasing concentrations (20, 200 nM) of Brig1 for 30 minutes at 37°C with or without subsequent treatment with SDS prior to electrophoresis. L, DNA size ladder.

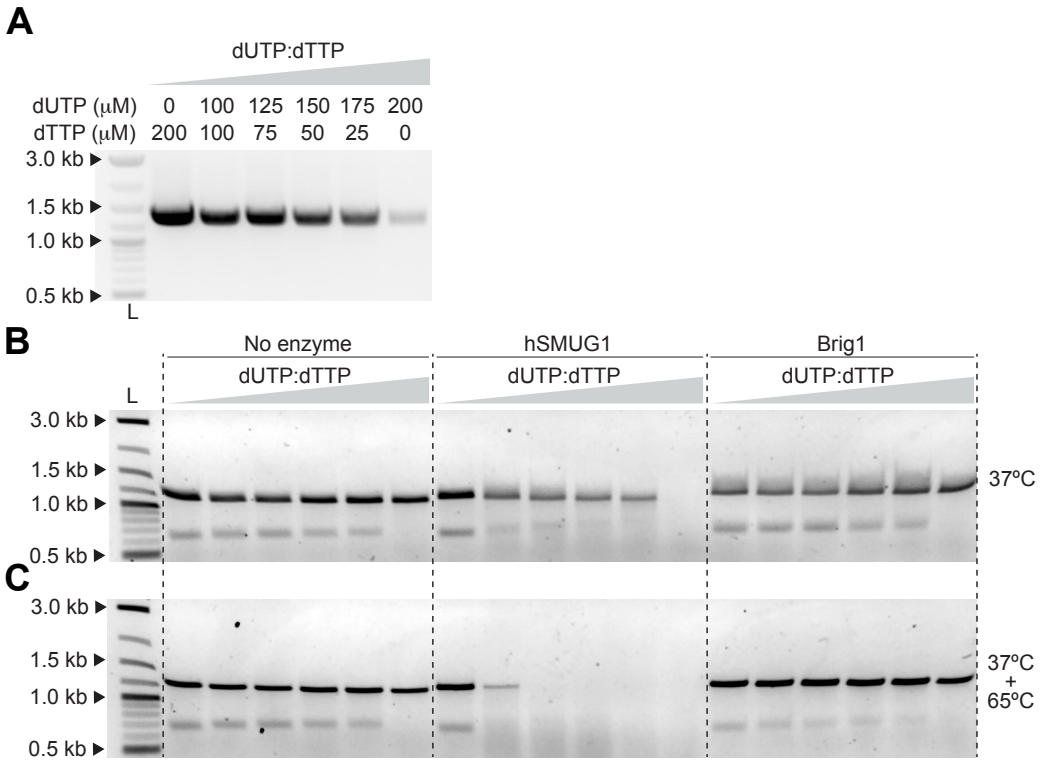


Figure 5.10 hSMUG1 degrades DNA containing a high concentration of uracil nucleobases in vitro

(**A**) Agarose gel electrophoresis of phage T4 *gp24* DNA PCR amplified with NEB Q5U DNA polymerase using increasing ratios of dUTP:dTTP to obtain PCR products with increasing concentrations of uracil in their DNA. (**B-C**) Agarose gel electrophoresis of gel purified PCR products from (**A**) (50 ng each) treated with 10 units of NEB hSMUG1 or 500 nM of Brig1 for 30 minutes at 37°C with (**C**) or without (**B**) heat treatment (20 minutes at 65°C) prior to electrophoresis. L, DNA size ladder.

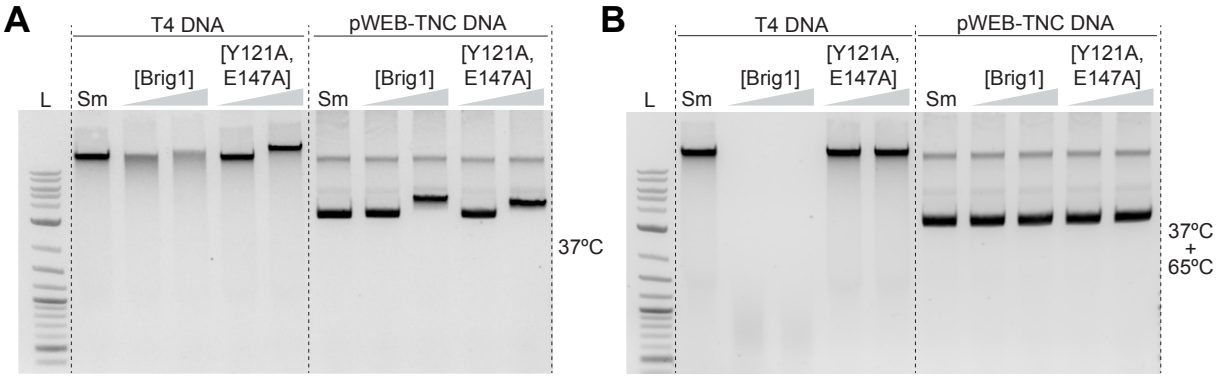


Figure 5.11 Mutations in the DNA glycosylase pocket of Brig1 abrogate in vitro T4 DNA degradation

(**A**) Agarose gel electrophoresis of T4 or pWEB-TNC DNA (500 ng) treated with increasing concentrations (50, 500 nM) of Brig1 or Brig1(Y121A, E147A), or with 10 units of NEB hSMUG1 (Sm), for 30 minutes at 37°C. (**B**) Same as (**A**) but heated for an additional 20 minutes at 65°C prior to gel electrophoresis. L, DNA size ladder.

5.8 Host DNA repair pathways do not play a specialized role in Brig1-mediated anti-phage activity

E. coli mechanisms of base excision repair could assist degradation at Brig1-generated abasic sites in vivo. In bacteria, bases that are damaged through deamination, oxidation or methylation are removed from the genome by DNA glycosylases to generate abasic sites that are then processed by host AP endonucleases. These endonucleases cleave the DNA backbone at the abasic site to generate a single-strand DNA nick to initiate repair (Krokan and Bjoras, 2013; Parikh et al., 1997). We wondered whether AP endonucleases are important for Brig1 immunity in vivo, possibly by nicking Brig1-generated abasic sites to accelerate the cleavage of viral DNA, as demonstrated in vitro (**Fig. 5.5-C**). To test this, we investigated Brig1 immunity against T4 in *E. coli* deletion mutants lacking either one or both of the two major *E. coli* AP endonucleases, Exonuclease III (XthA) and Endonuclease IV (Nfo) (Saporito and Cunningham, 1988; Saporito et al., 1988). We also tested immunity in hosts lacking the pyrimidine DNA glycosylase-lyase Endonuclease III (Nth) (Boiteux, 1993) or the abasic site sensor YedK (Paulin et al., 2022; Wang et al., 2019), other enzymes that participate in base excision repair in *E.*

coli. None of these deletions increased T4 PFU counts in the presence of Brig1 (**Fig. 5.12-A**), suggesting that the deleted genes are not individually required for immunity. We also performed the opposite experiment, i.e., overexpressing XthA and Nfo to determine if they enhance Brig1 immunity and found that neither of the AP endonucleases provided a further decrease in T4 PFUs (**Fig. 5.12-B**). Finally, we tested whether nucleases, helicases and recombinases involved in recombinational DNA repair – RecBCD, RecQ, RecJ and RecA (Dianov et al., 1994; Dillingham and Kowalczykowski, 2008; Heyer, 2004; Wright et al., 2018), which could process DNA ends generated by Brig1 DNA glycosylase and any subsequent host-encoded lyase activities – affect immunity. T4 PFUs formed in the presence of Brig1 did not change in any of the mutant hosts (**Fig. 5.12-C**), suggesting that these DNA repair pathways are not involved in Brig1 immunity. Although these results exclude the possibility that the major *E. coli* repair enzymes and AP endonucleases play a specialized role in Brig1-mediated anti-phage defense, we are unable to discount the known redundancy of host enzymes with AP lyase activity as well as the possibility of as-yet-unknown endonucleases, from bacteria or phage, that could act on Brig1-generated abasic sites in vivo to promote (or abrogate) phage DNA degradation or repair. Furthermore, abasic sites are not expected to persist in vivo given their highly reactive nature, resulting ultimately in a variety of DNA lesions including double-strand DNA breaks, interstrand DNA crosslinks and DNA-protein crosslinks (Thompson and Cortez, 2020). Whether and how these DNA lesions contribute to Brig1-mediated phage replication inhibition in vivo remains an open question.

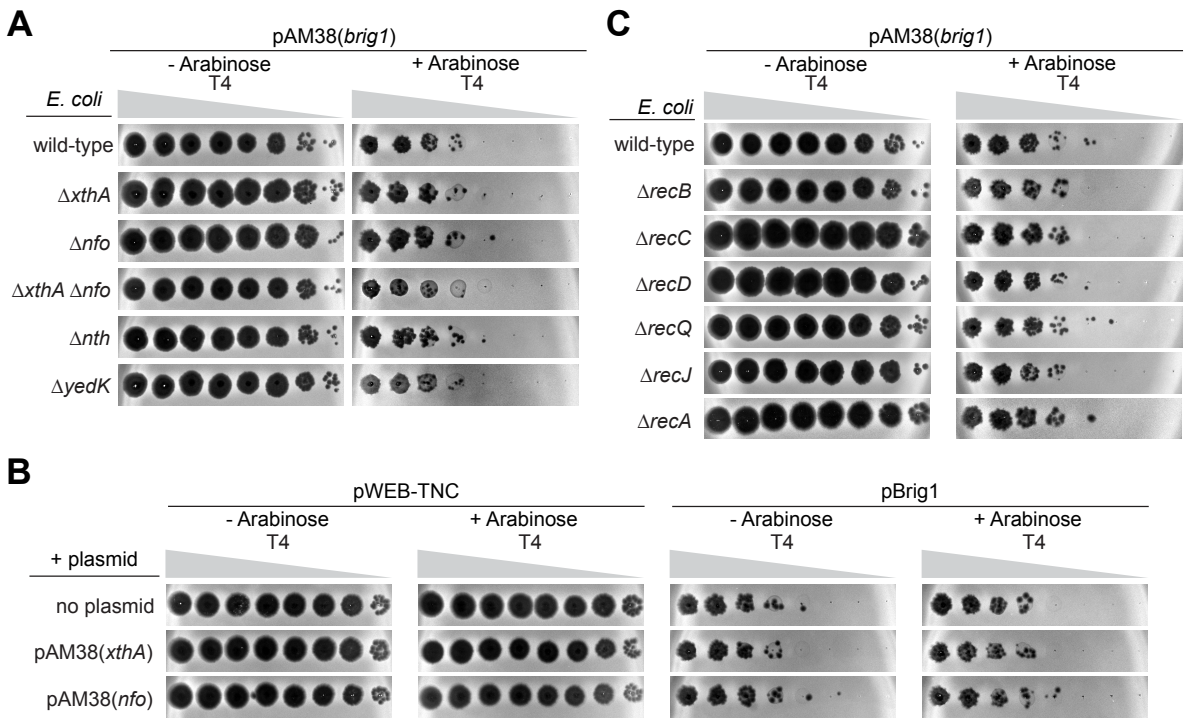


Figure 5.12 *E. coli* AP endonucleases and DNA repair proteins are not specifically required for Brig1 immunity

(**A**) Ten-fold serial dilutions of phage T4 on lawns of different *E. coli* K-12 BW25113 mutants with deletions of genes involved in base excision repair, carrying the pAM38 vector to express Brig1 using an arabinose-inducible promoter, in the presence (+ Arabinose) or absence (- Arabinose) of the inducer. (**B**) Ten-fold serial dilutions of phage T4 on lawns of *E. coli* EC100, each lawn carrying two plasmids: pWEB-TNC or pBrig1, and pAM38(*xthA*) or pAM38(*nfo*), which express the *E. coli* AP endonucleases XthA and Nfo, respectively, using an arabinose-inducible promoter, in the presence (+ Arabinose) or absence (- Arabinose) of the inducer. (**C**) Same as (**A**) but using *E. coli* K-12 BW25113 mutants with deletions of genes involved in RecABCD or RecJQ DNA repair pathways. Plaque images of one representative experiment from $n = 2$ independent experiments are shown in (**A-C**).

5.9 Summary

In this chapter, I describe a thorough biochemical characterization of Brig1, the prokaryotic DNA glycosylase we discovered that provides immunity in *E. coli* against bacteriophage T4. Using AlphaFold2 structural analyses, in vitro electrophoresis-

based assays, and mass spectrometry, we showed that Brig1 generates abasic sites in both ssDNA and dsDNA containing target alpha-glucosyl-hmC nucleobases. Our assays demonstrated a lack of substantial DNA strand cleavage activity by Brig1 following base excision, suggesting that Brig1 is a monofunctional DNA glycosylase that lacks AP lyase activity. We demonstrated that the enzyme's base excision activity can be abrogated through mutation of the DNA glycosylase pocket. Finally, we found that base excision activity mediated by the DNA glycosylase pocket of Brig1 resulted in T4 phage DNA degradation in vitro.

In summary, Brig1 is distinct from restriction enzymes that cleave DNA to generate double-strand DNA breaks. Brig1 represents a new family of DNA glycosylases, related to the uracil DNA glycosylase superfamily, with an enzymatic pocket that can accommodate a bulky glucosylated pyrimidine nucleobase. In the next chapter, I will investigate Brig1-mediated immunity against diverse T-even phages and look at Brig1 homologs in the context of bacterial defense islands.

CHAPTER 6. BRIG1 AND ITS HOMOLOGS PROVIDE IMMUNITY AGAINST DIVERSE T-EVEN BACTERIOPHAGES

6.1 Background

In Chapters 4 and 5, I describe the discovery and characterization of Brig1, a novel DNA glycosylase, related to the uracil DNA glycosylase superfamily, which provides antiviral defense against phage T4 in *E. coli*. Brig1 specifically excises alpha-glucosyl-hydroxymethylcytosine nucleobases in the phage DNA, generating abasic sites to inhibit phage DNA replication. In this chapter, I will first describe this enzyme's ability to provide antiviral defense against diverse T-even phages, the subfamily of myoviruses that T4 belongs to. Finally, I will look at homologs of Brig1 and their roles in anti-phage defense.

6.2 Brig1 provides immunity against T-even phages that carry alpha-glucosyl-hmC nucleobases

To test the range of phages restricted by Brig1, we infected *E. coli* with seven different coliphages and found that, in addition to T4, phages T2 and T6 were highly sensitive to Brig1 targeting (**Fig. 6.1**). This is most likely due to the presence of alpha-glucosyl-hmC nucleobases in T-even phage genomes (Lehman and Pratt, 1960). Indeed, 70% of hmC sites are alpha-glucosylated in phage T2 (similar to T4), but only 3% are in the case of T6 (Kuno and Lehman, 1962; Lehman and Pratt, 1960). Unlike T4, neither T2 nor T6 harbor beta-glucosylated hmC, but instead carry

beta-1,6-glucosyl-alpha-glucose (gentiobiose) adducts on 5% and 72% of T2 and T6's hmC nucleobases, respectively (**Fig. 6.2-A**). Since the majority of the hmC nucleobases in the T2 genome are alpha-glucosylated, this phage is expectedly very sensitive to Brig1 targeting (**Fig. 6.1**). On the other hand, since only a small fraction of the T6 genome contains alpha-glucosyl-hmC, the high susceptibility of this phage to Brig1 (**Fig. 6.1**) is intriguing. To investigate this, we isolated two T6 phages that escaped targeting (**Fig. 6.2-B**) and found that both carried inactivating mutations in the T6 *a-gt* gene (**Table 6.1**). Expression of phage T4 *a-gt* re-sensitized both T6 escapers to Brig1 immunity (**Fig. 6.2-B**), a result that demonstrates the requirement of alpha-glucosylated hmC nucleobases in the T6 genome for immunity against this phage.

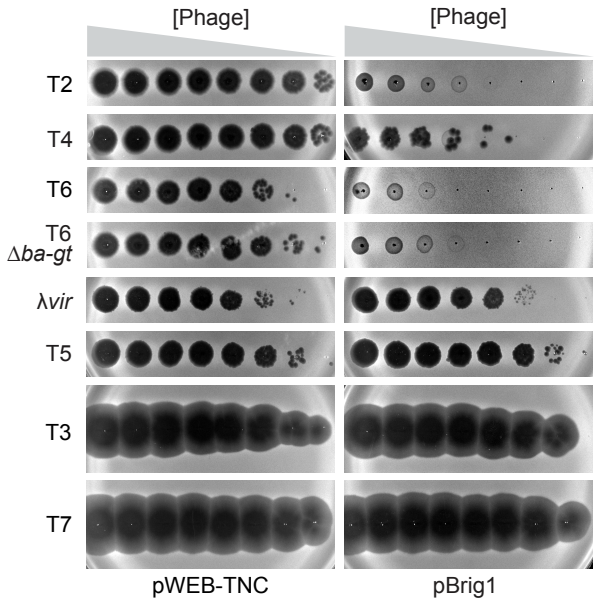


Figure 6.1 Brig1 provides immunity against T-even coliphages

Ten-fold serial dilutions of common coliphages spotted on lawns of *E. coli* EC100 carrying cosmid pWEB-TNC or pBrig1. T6 Δ*ba-gt* lacks the glucosyltransferase that adds the second glucose to alpha-glucosyl-hydroxymethylcytosine nucleobases in phage T6. Plaque images of one representative experiment from $n = 3$ independent experiments are shown.

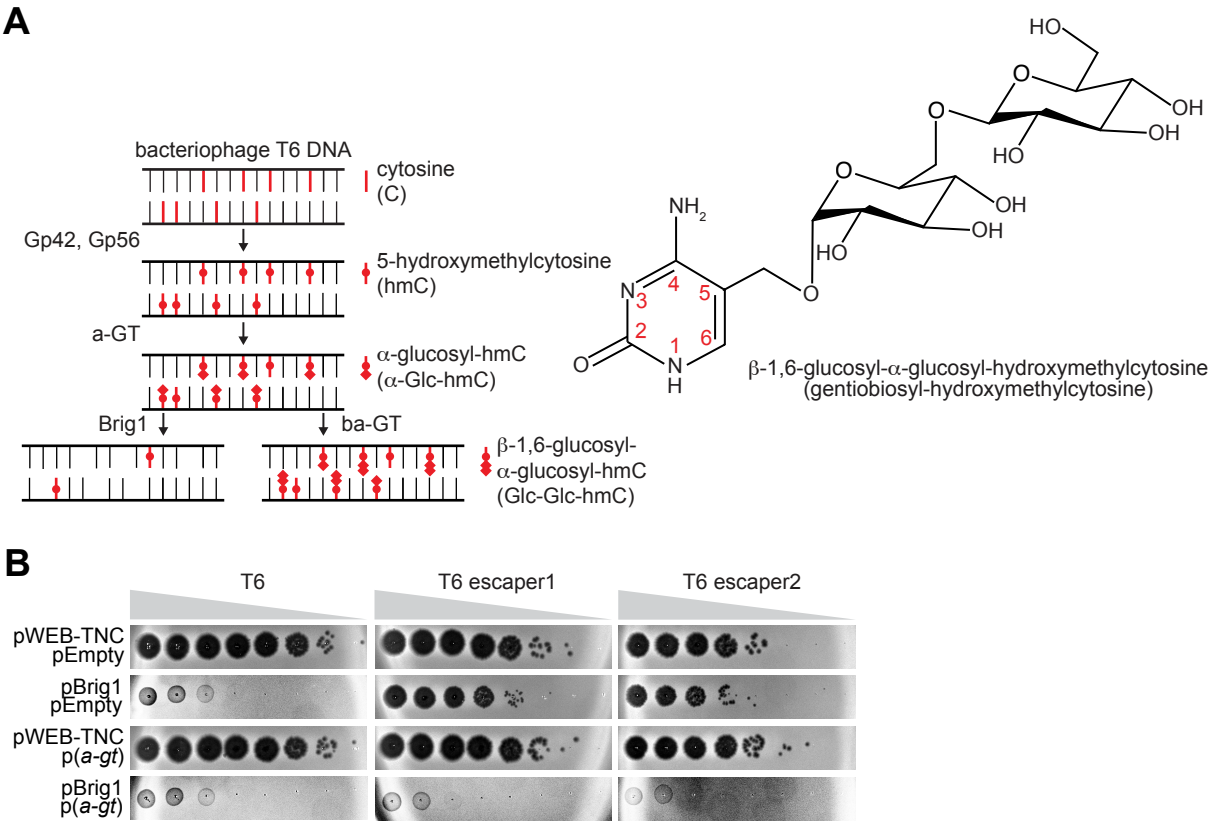


Figure 6.2 Alpha-glucosylation of hydroxymethylcytosine nucleobases sensitizes T6 phage to Brig1 immunity

(**A**) Schematic of the cytosine modification pathway in phage T6, including a model for Brig1 immunity in which the DNA glycosylase recognizes and excises alpha-glucosyl-hydroxymethylcytosine nucleobases from the viral genome, before the addition of the second glucosyl group to generate gentiobiosyl-hydroxymethylcytosine. T6 enzymes: Gp42, dCMP hydroxymethylase; Gp56, dCTPase; a-GT, alpha-glucosyltransferase; ba-GT, beta-alpha glucosyltransferase. (**B**) Ten-fold serial dilutions of T6, T6 escaper1 and T6 escaper2 phages on lawns of *E. coli* EC100, each lawn carrying two plasmids: pWEB-TNC or pBrig1, and pEmpty or p(*a-gt*).

Table 6.1 Mutations in the *a-gt* gene of T6 phages that escape Brig1 immunity

escaper	Mutations in *a-gt* gene
T6-1	nt 716 C>T mutation (Ala239->Val239)
T6-2	1 bp insertion (C) after nt 758 (frameshift)

6.3 Alternative glycosylation of hydroxymethylcytosine nucleobases in T-even phages abrogates Brig1 activity in vivo and in vitro

Next, we extracted T2, T4 and T6 phage DNA and treated them with purified Brig1 (**Fig. 6.3**). As with phage T4 DNA, T2 DNA was partially degraded at 37°C and not detected after subsequent treatment at 65°C. In contrast, T6 DNA was not visibly degraded by Brig1, even after heat treatment. We hypothesized that this could be due to the low fraction of alpha-glucosyl-hmC nucleobases in the T6 genome. To test this, we constructed a T6 phage lacking the *ba-gt* gene, which encodes beta-alpha glucosyltransferase (ba-GT), the enzyme required to add the second glucose in beta-linkage to alpha-glucosyl-hmC nucleobases and generate gentiobiosyl-hmC (**Fig. 6.2-A**). This phage, T6 Δ*ba-gt*, only carries alpha-glucosylated hmC nucleobases (presumably 75% of the cytosines in T6, **Fig. 6.2-A**), and is more susceptible to Brig1 immunity than wild-type T6 (**Fig. 6.4-A**). In addition, treatment of T6 Δ*ba-gt* DNA with Brig1 resulted in complete degradation (**Fig. 6.4-B**). These results suggest that while gentiobiose modifications render T6 DNA resistant to Brig1 in vitro and most likely also in vivo, there is a window during the viral lytic cycle, after the activity of a-GT on newly replicated hmC nucleobases (Lehman and Pratt, 1960; Sommer et al., 2004) but before the addition of the second glucose by ba-GT, in which a large proportion of the hmC nucleobases in T6 are modified only with alpha-glucose and therefore susceptible to Brig1 restriction. Therefore, while additional modification of alpha-glucosyl-hmC nucleobases prevent Brig1 activity, their transient presence during the viral lytic cycle is sufficient for efficient immunity.

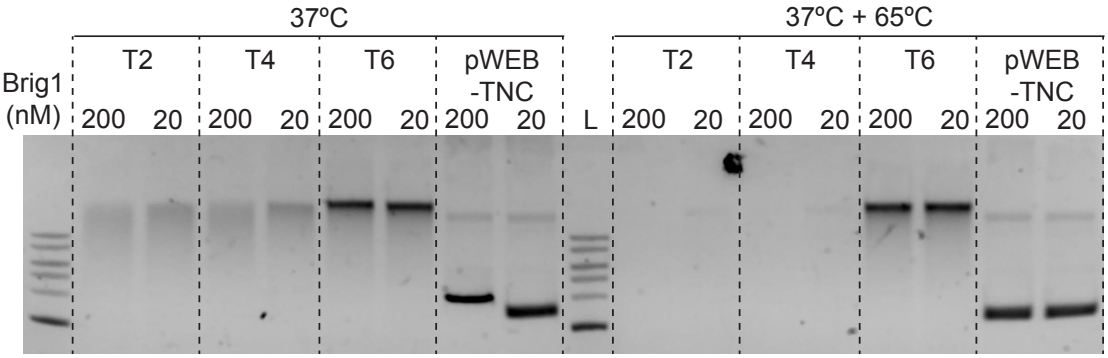


Figure 6.3 Brig1 degrades T2 and T4, but not T6, phage DNA in vitro

Agarose gel electrophoresis of T2, T4, T6 or pWEB-TNC DNA (125 ng) treated with decreasing concentrations (200, 20 nM) of Brig1 for 30 minutes at 37°C with or without heat treatment (20 minutes at 65°C) prior to electrophoresis. L, DNA size ladder.

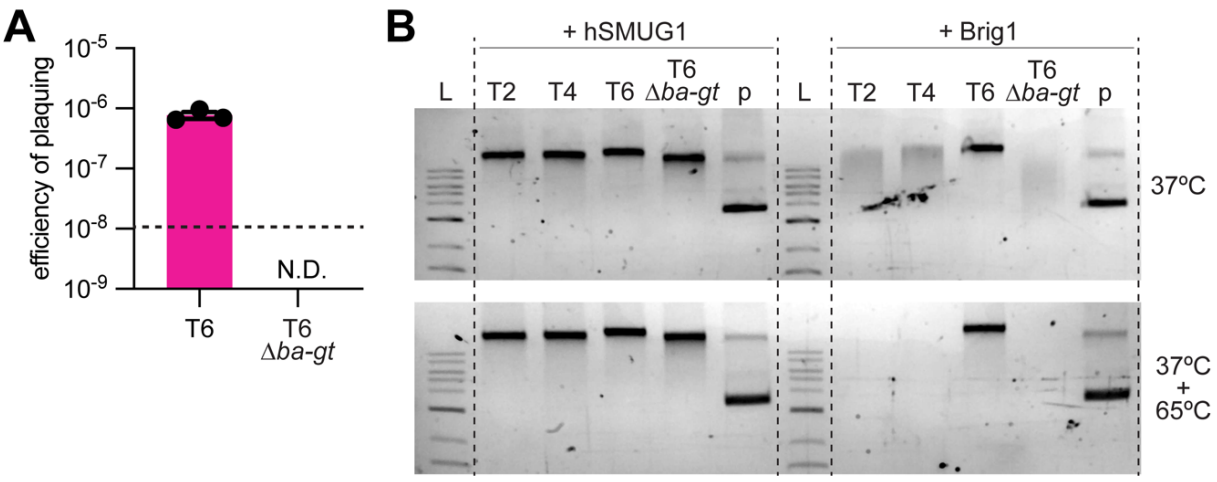


Figure 6.4 Gentiobiose modifications protect T6 phage DNA against Brig1-mediated base excision

(**A**) Efficiency of plaquing of T6 and T6 Δ*ba-gt* phages on lawns of *E. coli* EC100 carrying pBrig1. Mean ± SEM values are reported for $n = 3$ independent experiments. N.D., no plaques detected; dotted line, limit of detection. (**B**) Agarose gel electrophoresis of T2, T4, T6, T6 Δ*ba-gt* or pWEB-TNC DNA (50 ng) treated with 10 units of NEB hSMUG1 or 100 nM of Brig1 for 30 minutes at 37°C with (bottom) or without (top) heat treatment (20 minutes at 65°C) prior to electrophoresis. L, DNA size ladder.

Finally, we passaged T4 phage on an *E. coli* strain that overexpressed the T4 enzyme b-GT (beta-glucosyltransferase). T4 b-GT adds glucose in beta-linkage to ~30% of the hmC residues in the T4 genome during a typical viral lytic cycle in the absence of artificial overexpression (Lehman and Pratt, 1960; Sommer et al., 2004). Our goal was to investigate the effect of an increase in the fraction of beta-glucosylated hmC nucleobases in T4 genomic DNA on Brig1 activity. We found that the DNA from this phage, T4(+b-GT), was less susceptible to Brig1 degradation in vitro compared to wild-type phage (**Fig. 6.5**), since the DNA from T4(+b-GT) contains a lower fraction of alpha-glucosyl-hmC nucleobases. Altogether these data suggest that differential hypermodification of hmC nucleobases can serve as a viral strategy to counteract Brig1 targeting.

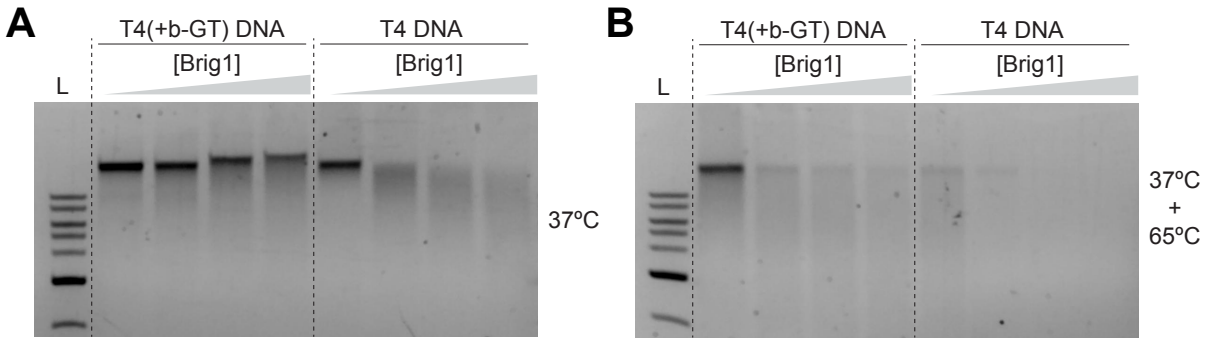


Figure 6.5 A higher fraction of beta-glucosyl-hmC nucleobases in T4 genomic DNA abrogates Brig1-mediated DNA degradation

(**A**) Agarose gel electrophoresis of Brig1-treated DNA (125 ng) from T4 phage passaged through *E. coli* EC100 or *E. coli* EC100/p(*b-gt*) which overexpresses T4 beta-glucosyltransferase to increase the frequency of beta-glucosyl-hydroxymethylcytosine modifications within the T4 genome [T4(+b-GT)]. DNA was incubated with Brig1 (2, 20, 200, 400 nM) for 30 minutes at 37°C prior to electrophoresis. L, DNA size ladder. (**B**) Same as (**A**) but heated for an additional 20 minutes at 65°C prior to gel electrophoresis.

6.4 Diverse T-even phages carry alpha-glucosyl-hmC nucleobases and are susceptible to Brig1 targeting

We next tested 69 different *E. coli* phages from the BASEL collection (Maffei et al., 2021) and found that Brig1 provides immunity against Bas35-45, all members of the T-even family that modify their genomes with alpha-glucosyl-hmC (**Fig. 6.6-A**). In contrast, plaque formation by two other T-even phages within the collection, Bas46-47, predicted to carry arabinosyl-hmC nucleobases instead of glucosyl-hmC (Maffei et al., 2021; Thomas et al., 2018), was not affected by Brig1 (**Fig. 6.6-B**). Overall, these data further confirm that Brig1 restricts T-even phages that contain alpha-glucosylated hmC residues. Furthermore, the high representation of T-even phages susceptible to Brig1 targeting suggests that alpha-glucosyl-hmC nucleobases are widespread among T-even myophages.

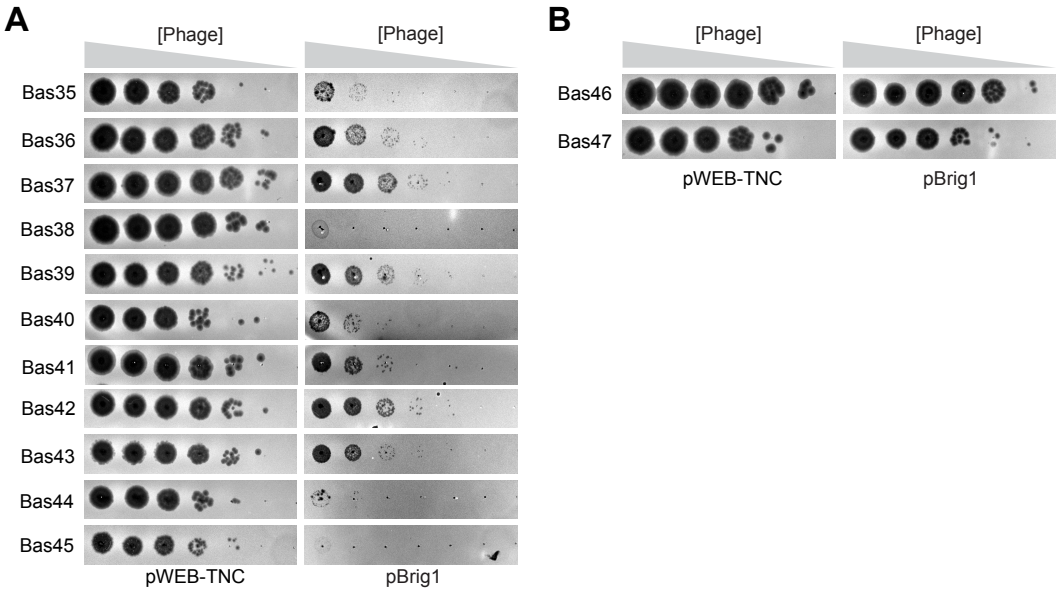


Figure 6.6 Diverse T-even phages are susceptible to Brig1 targeting

(**A**) Ten-fold serial dilutions of phages from the BASEL collection Bas35-45 spotted on lawns of *E. coli* EC100 carrying pWEB-TNC or pBrig1. (**B**) Same as (**A**) but with Bas 46-47 phages from the BASEL collection. Plaque images of one representative experiment from $n = 3$ independent experiments are shown in (**A-B**).

6.5 Homologs of Brig1 located within prokaryotic defense islands provide immunity against T-even phages

We used PSI-BLAST to analyze the prevalence of Brig1 in prokaryotic genomes and found 42 non-redundant homologs (**Fig. 6.7**), all annotated on NCBI databases as hypothetical proteins. Many of these are present within putative phage defense islands (**Fig. 6.8**), near other annotated phage defense genes. Furthermore, many of the putative Brig1 homologs are associated with phage growth limitation (Pgl) genes found in Pgl and BREX systems (Goldfarb et al., 2015; Hoskisson et al., 2015). Interestingly, BREX systems also rely on differential modification of host and viral DNA to recognize their targets (Goldfarb et al., 2015; Gordeeva et al., 2018). Most of the Brig1 homologs currently available in genetic databases are found in Actinobacteria (**Fig. 6.7**).

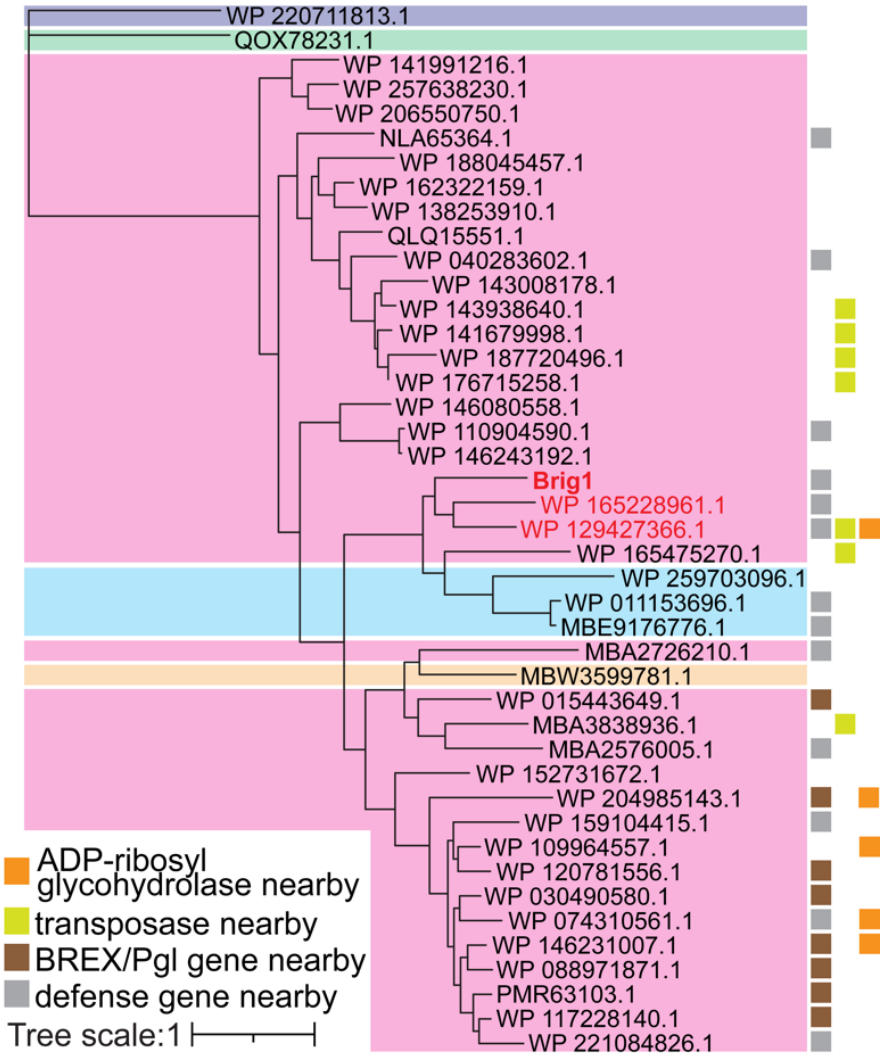


Figure 6.7 Phylogenetic tree of Brig1 homologs

Maximum likelihood tree of 42 Brig1 homologs (noted by their NCBI protein accession numbers) found in different phyla: Firmicutes (purple background), Proteobacteria (green), Actinobacteria (pink), Cyanobacteria (light blue) and Planctomycetota (orange). Brig1 homologs that provide phage defense against T4 and T6 are indicated in red. Grey squares indicate the presence of putative defense genes in the immediate vicinity (within 10 genes upstream and downstream); brown squares, BREX/Pgl genes; yellow squares, transposases; orange squares, ADP-ribosyl glycohydrolases.

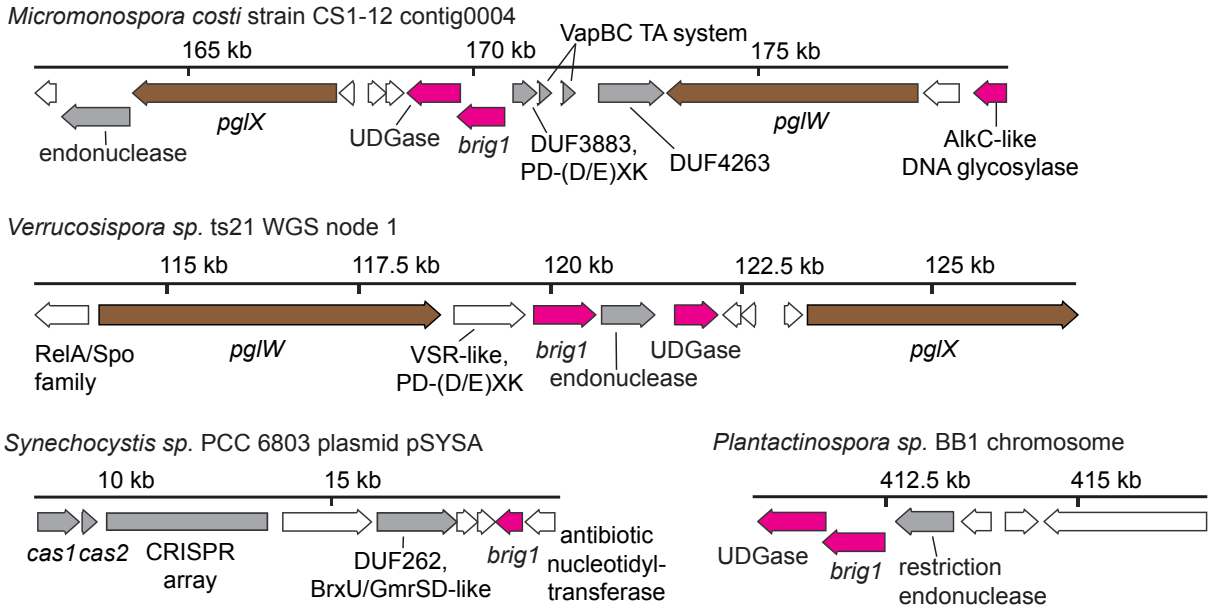


Figure 6.8 Brig1 homologs in bacterial defense islands

Gene neighborhoods of Brig1 homologs found in putative phage defense islands. Brig1 homologs and other DNA glycosylases are shown in magenta, Pgl/BREX genes in brown, and other putative antiviral genes in grey.

We tested different Brig1 homologs for their ability to defend *E. coli* against T4 and T6 phage infection using plaque assays. We found that two closely related Brig1 homologs, both present in *Nocardioides*, provided protection (**Fig. 6.7** and **Fig. 6.9-A**). These homologs are present in putative defense islands (**Fig. 6.9-BC**), with the one harbored by *Nocardioides zhouii* located in a similar genomic neighborhood as Brig1, i.e., adjacent to a predicted ADP-ribosyl glycohydrolase and near a ThsA-like, SIR2-domain protein (Fig. 6c). Both share ~50% amino acid identity with Brig1 and a high level of predicted structural similarity (**Fig. 6.10**). Interestingly, the *N. anomalus* homolog contains an additional C-terminal domain of unknown function (**Fig. 6.10-F**) that is also present in some ADP-ribosyl glycohydrolases.

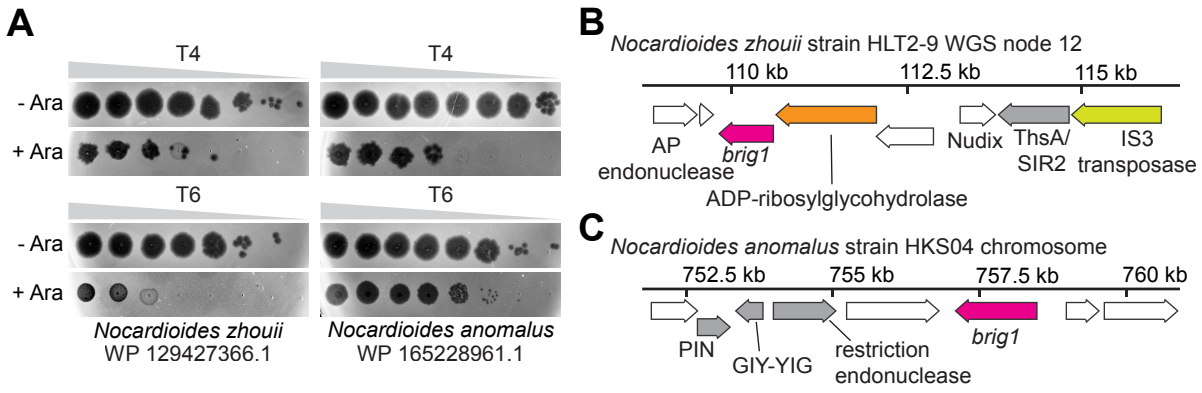


Figure 6.9 Brig1 homologs provide defense against T-even phages

(**A**) Ten-fold serial dilutions of phage T4 or T6 on lawns of *E. coli* EC100 carrying the pAM39 vector to express the indicated Brig1 homologs using an arabinose-inducible promoter, in the presence (+ Ara) or absence (- Ara) of the inducer. Plaque images of one representative experiment from $n = 3$ independent experiments are shown. (**B**) Gene neighborhood of the Brig1 homolog from *Nocardioides zhouii* showing neighboring ADP-ribosyl glycohydrolase, transposase and putative phage defense genes in orange, yellow and grey, respectively. (**C**) Same as (**B**) but for the Brig1 homolog from *Nocardioides anomalus*.

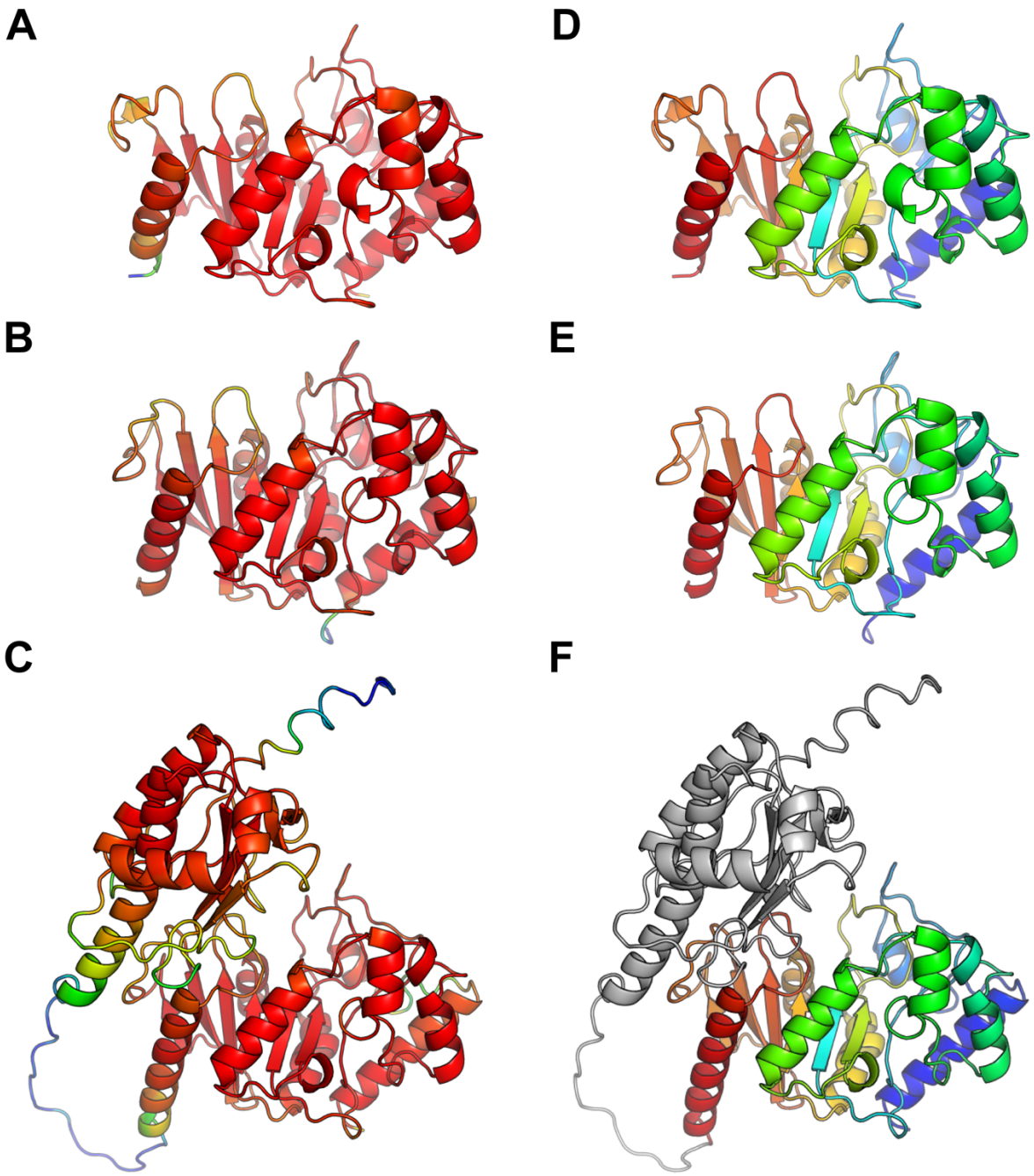


Figure 6.10 AlphaFold2 structures of Brig1 homologs

(**A-C**) AlphaFold2 structures of Brig1 (**A**) and its homologs from *Nocardioides zhouii* (**B**) and *Nocardioides anomalus* (**C**), colored by b-factors. Red to blue spectrum represents high to low confidence of secondary structure prediction. (**D-F**) AlphaFold2 structures of Brig1 (**D**) and its homologs from *Nocardioides zhouii* (**E**) and *Nocardioides anomalus* (**F**), colored by N- to C-terminal (blue to red). A C-terminal domain of unknown function is shown in grey in (**F**).

6.6 Summary

In summary, bacteria adapt to phage predation through the evolution of a vast assortment of defense systems. Here, we report the discovery of a novel mechanism of anti-phage defense through a DNA repair protein, a DNA glycosylase, that appears to be specifically co-opted by prokaryotes for this purpose. This discovery came about through infecting *E. coli* carrying a soil metagenomic DNA library with the lytic coliphage T4 to isolate clones carrying protective genes. Following this approach, we identified Brig1, a DNA glycosylase that excises alpha-glucosyl-hmC nucleobases from the bacteriophage T4 genome to generate abasic sites and inhibit viral replication. Using biochemistry approaches, we showed that Brig1 generates abasic sites in both ssDNA and dsDNA carrying target nucleobases. Furthermore, Brig1-mediated base excision is not followed by subsequent cleavage of DNA strands, thereby distinguishing its mechanism of action from classical restriction enzymes that generate double-strand DNA breaks.

We found that Brig1 provides defense against diverse T-even phages given that they carry alpha-glucosyl-hmC nucleobases in their genomes. Brig1 homologs that provide immunity against T-even phages are present in multiple phage defense loci across distinct clades of bacteria. Interestingly, Brig1 is related to the superfamily of uracil DNA glycosylases and the large glycosylase pocket of Brig1, to accommodate a bulkly glucosylated pyrimidine, expands the known evolutionary potential of this enzymatic scaffold. Our study highlights the benefits of screening unsequenced DNA

and reveals prokaryotic DNA glycosylases as important players in the bacteria-phage arms race.

CHAPTER 7. DISCUSSION

7.1 Bacteriophage recombination systems limit CRISPR-Cas targeting through the generation of escape mutations

In Chapters 2 and 3, I investigate how bacteriophage λ escapes DNA cleavage by type I and II CRISPR-Cas nucleases (Hossain et al., 2021). We showed that the λ *red* operon enables the accumulation of escaper phage, and thus increases the propagation of the phage by several orders of magnitude. We propose a model in which this phenomenon is a result of the function of the λ Red system in the repair of viral DNA cleaved by RNA-guided Cas nucleases: Gam prevents both the rapid degradation of phage DNA by RecBCD and repair through *E. coli* RecA-mediated recombination at *chi*-like sequences, leaving the DSBs generated by CRISPR cleavage available for repair by Exo-Beta recombination (**Fig. 7.1**). Exo-Beta repair then leads to the accumulation of a greater number of escaper phages harboring distinct point mutations in the CRISPR target sequence. We found that Exo-Beta also mediates other forms of repair, in which recombination occurs with a homologous sequence in the host chromosome (**Fig. 2.3**, **Fig. 3.10-3.11** and **Table 3.14-3.15**), or between short homologous sequences flanking the target site in the viral genome (**Fig. 3.12** and **Table 3.16**). However, we believe that these examples represent rare cases, since they require either the presence of phage-related chromosomal sequences or the generation of non-deleterious deletions (less common within compact phage genomes (Brüssow and Hendrix, 2002)), respectively.

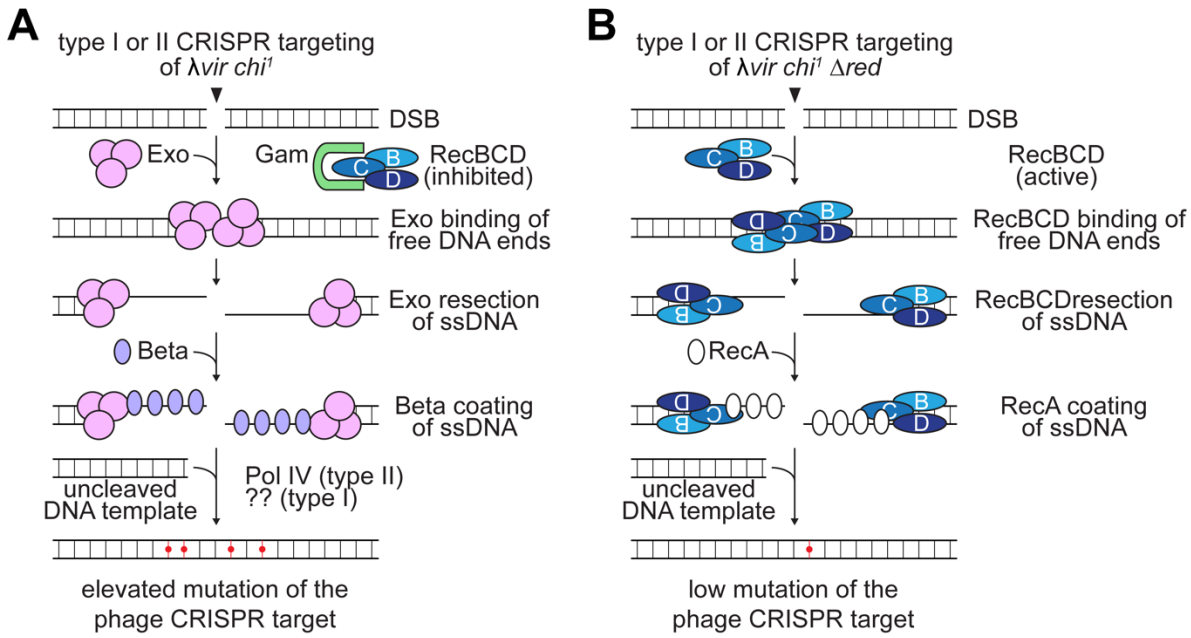


Figure 7.1 λ Red recombination mediates evasion of CRISPR-Cas targeting

(**A**) Model for Red-mediated repair of λ phage DNA cleaved by CRISPR nucleases. The DNA ends at the dsDNA break (DSB) generated by the action of Cas9 or Cascade-Cas3 nucleases are resected by Exo, generating 3' overhangs that are bound by Beta. At the same time Gam prevents access of the host's RecBCD complex to the free DNA ends. Finally, Beta promotes recombination with an intact copy of λ DNA. Through an unknown mechanism that involves Pol IV for Cas9-cleaved DNA, a relatively high number of mutations are introduced at the repaired target site, which, if they enable escape from CRISPR targeting, are selected and spread through the viral population. When more intact copies of the phage DNA are present in the cell, more phage escapers are generated. (**B**) Model for the repair of λ phage DNA cleaved by CRISPR nucleases in the absence of Red. In the absence of the phage recombination system, the most common pathway for the repair of DNA breaks in *E. coli* is RecABCD. RecBCD is the exonuclease that generates 3' overhangs, which, upon encountering a *chi* site, are bound by RecA to promote recombination with an intact copy of the phage DNA. This repair pathway generates a significantly lower number of phage escapers than Exo-Beta.

Most repair scenarios will require an intact copy of the phage target DNA for recombination to re-generate the original sequence. Consistent with this, we found that the Red system is particularly efficient at increasing the number of escapers in conditions of weaker defense, when a considerable fraction of wild-type phage will

remain uncleaved (**Fig. 7.2**). Given that only specific seed and PAM target mutations prevent effective CRISPR-Cas target recognition and/or cleavage, multiple rounds of cleavage and repair are probably necessary before *de novo* escape mutations become prevalent in the viral population. Indeed, from our liquid culture time course experiment, during the first 9 hours of infection of cells expressing Cas9 and the weak-targeting *spc45c* crRNA (and presumably *spc9* and *spc40* crRNAs as well) most phages contained intact, wild-type target sequences (**Fig. 3.9-B** and **Table 3.12**), suggesting that the Exo-Beta system was able to efficiently and accurately repair the DSBs generated during type II-A targeting, allowing the phage to replicate and continue its lytic cycle. Phages carrying target mutations only become detectable within the population relatively late during infection (**Table 3.12**). Although the type II-A spacers evaluated in this study were not naturally acquired during the CRISPR response to λ infection, a previous study from the Marraffini lab has shown that weak spacers that cannot provide high fitness to the host are not only acquired but also maintained in the bacterial population long after infection (Heler et al., 2019). While these findings were obtained investigating the type II-A CRISPR-Cas immune response to the staphylococcal phage ΦNM4Φ4 (a λ-like phage also belonging to the *Siphoviridae* family), we presume that weak spacers will also be abundant in hosts harboring this CRISPR system following infection with other phages. In the specific case of phage λ, of the 4 spacers acquired by the native *E. coli* type I-E system (Strotskaya et al., 2017) that we tested in this study, we found one that provided weak defense. We believe that the relative abundance of spacers that mediate poor

targeting could be important to increase the targeting diversity, which helps prevent the rise of escaper phages (van Houte et al., 2016). In addition, phage-encoded anti-CRISPRs can diminish the effectiveness of strong spacers (Borges et al., 2018; Landsberger et al., 2018). For all these reasons, we believe that the evasion of weak targeting bestowed by Red-like systems could have considerable effects for the phage-CRISPR arms race.

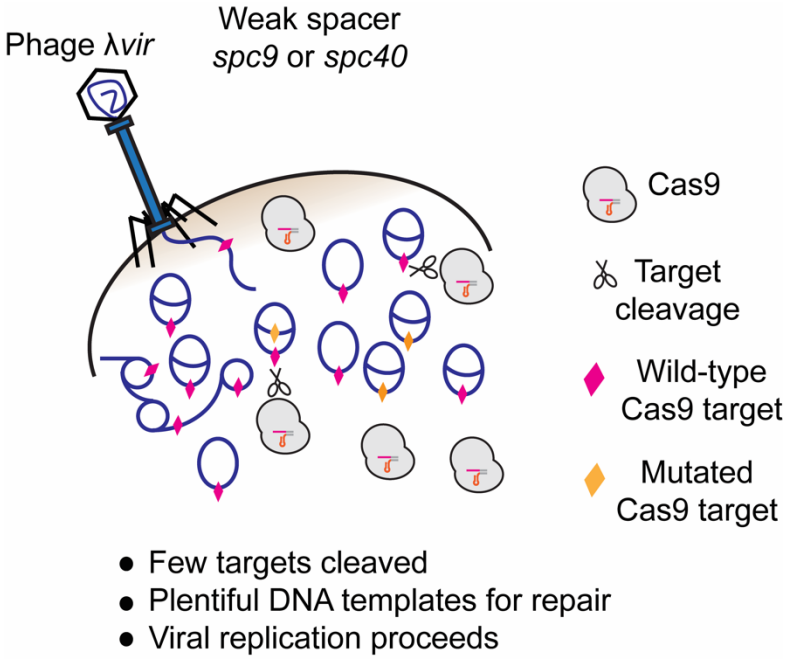


Figure 7.2 λ Red recombination is enabled in the presence of uncleaved phage DNA repair templates

Red-mediated repair of λ phage DNA cleaved by Cas9 requires a large presence of uncleaved phage DNA target sequences, which can serve as templates for homologous recombination. We hypothesize that this scenario occurs with “weak” spacers, for example *spc9* and *spc40* in our study, where a large fraction of escaper generation can be attributed to the presence of λ Red. In these scenarios, we see strong evasion of CRISPR-Cas targeting and higher levels of phage survival.

When the CRISPR defense is strong and presumably most copies of the infecting DNA are rapidly cleaved, we obtained different results for Cas9 and Cascade-Cas3 targeting. In the experiment where Cas9 was programmed with the strong spacer *spc45*, the number of escaper plaques was very low, both in the presence and absence of Red (**Fig. 2.8** and **Fig. 2.9-C**). We believe that rapid cleavage of a large fraction of the phage DNA leaves very few substrates for Exo-Beta recombination and repair (**Fig 3.7-B** and **Fig. 7.3**). We conclude that in this experiment the phages that evade Cas9(*spc45*) targeting most likely contain pre-existing mutations. In support of this hypothesis, the frequency of escapers for *spc45* was found to be approximately 10^{-6} (**Fig. 2.9-C** and **Fig. 3.7-C**; with the exception of an outlier datapoint in **Fig. 2.9-C**), which is close to the value expected from the mutation frequency previously determined for phage λ, 7.7×10^{-8} mutations per base pair (Drake, 1991), considering that about 10 base pairs of the target sequence (seed or PAM) can be mutated to avoid Cas9 cleavage. In contrast, in competition experiments in which wild-type and Δ*red* phages were used to co-infect the Cas9(*spc45*) host, all 36 plaques tested contained λ phages harboring *red* (**Fig. 2.17**) and target mutations with a consistent mutation pattern (**Table 3.4**). This suggests that, even during strong targeting conditions, some uncleaved template is still available for Exo-Beta recombination and repair, especially during infections at high MOI, which can lead to the introduction of mutations and their spread during a more sensitive assay such as competition. On the other hand, escape from targeting mediated by Cascade-Cas3 programmed with spacers that provide strong defense

was substantially increased in the presence of Red (**Fig. 3.15**). We attribute these different results to the contrasting DNA degradation activities of each of these nucleases and hypothesize that Exo-Beta can more efficiently recombine the products of Cas3 ssDNA cleavage and processive degradation than the blunt dsDNA ends generated by Cas9.

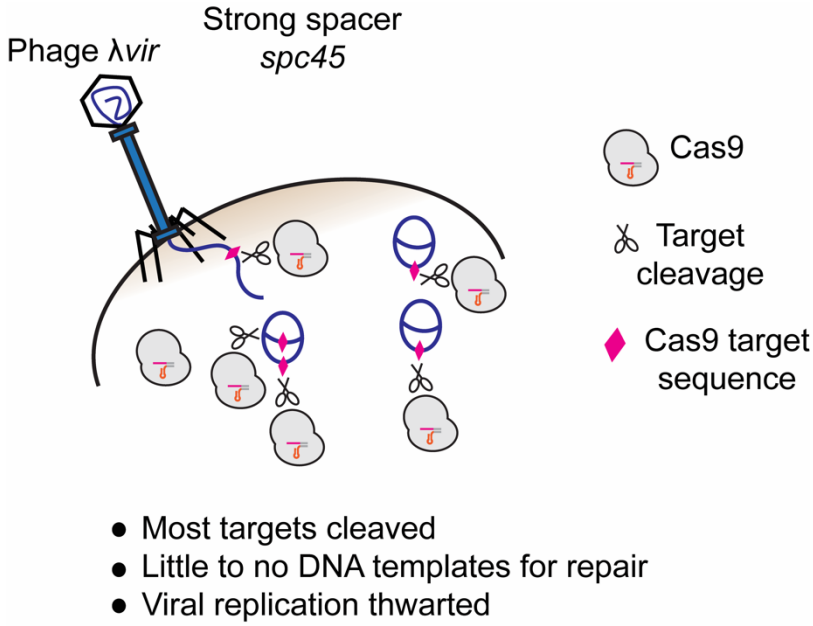


Figure 7.3 λ Red recombination is blunted when there is a paucity of uncleaved phage DNA repair templates

Red-mediated repair of λ phage DNA is inefficient when there are very few uncleaved target sequences, meaning limited templates for homologous recombination. Template matching is time consuming and less successful. We hypothesize that this scenario occurs with “strong” spacers, for example *spc45* in our study, which rapidly cleave phage DNA during infection and where λ Red is unable to mediate any meaningful evasion of CRISPR-Cas targeting. In these scenarios, we see limited phage survival during Cas9 targeting.

Regardless of the type of escaper we observed (**Fig. 2.6**), our data showed a substantial increase in the number of escapers where Exo-Beta but not RecABCD

recombination is functional (wild-type, *red*$^{+}$ phages) when compared to infection in conditions where RecABCD but not Exo-Beta repair is active (*chi*^{1} Δ*red* phages). Given that Cas9 remains bound to the cut DNA ends after cleavage (Garneau et al., 2010; Sternberg et al., 2014), one simple explanation for our results could be that while Exo effectively displaces the Cas9 nuclease to initiate recombination, RecBCD cannot do this efficiently, reducing the number of genomes that it can repair. However, during infection with Δ*gam* phages, when both repair pathways are functional, the number of escapers is similar to those obtained for the Δ*red* phage (**Fig. 2.9** and **Fig. 3.14**). This result suggests that RecBCD can capture and degrade the free DNA ends more efficiently than Exo. In addition, in the absence of Gam but with expression of Exo-Beta, the host RecBCD nuclease was able to extensively degrade the viral DNA (**Fig. 2.13**), a result that also argues against the possibility that Cas9, and even Exo, blocks RecBCD access to free DNA ends. Another factor that could affect the differences in escaper generation between the two repair systems is the different sequence requirements for recombination. While Exo-Beta activity is constitutive and does not depend on specific sequences, RecBCD only initiates recombination upon encountering a *chi* site, and even then there is only a 40% chance that the complex will successfully switch modes from degradation to repair (Taylor and Smith, 1992). Therefore, even in the presence of multiple *chi* sites, the efficiency of recombination of RecBCD, and hence the generation of escaper mutations, is lower than the constitutive Exo-Beta (**Fig. 2.15**). However, even in the presence of “constitutive” RecABC recombination (in the host with the *recD* deletion),

escaper generation is improved but still lags λ Exo-Beta (**Fig. 2.16**), a result which suggests that the DNA repair mediated by Exo-Beta recombination provides a specialized route for phage escape (**Fig. 7.4**).

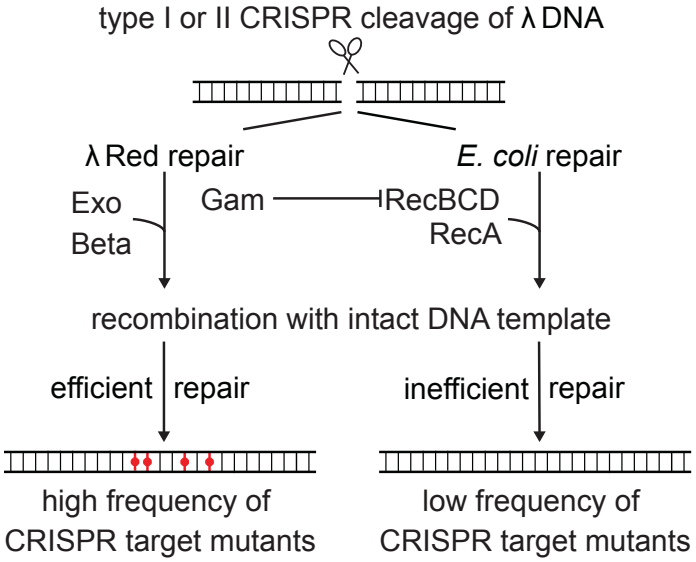


Figure 7.4 λ Red recombination provides a specialized route for phage escape from CRISPR-Cas immunity

Red-mediated recombination of λ phage DNA is more efficient than the host's RecBCD-RecA recombination pathway in repairing Cas9-cleaved phage DNA, even in the presence of constitutive host recombination.

Exo-Beta repair preferentially enables at least three distinct types of escapers (**Fig. 2.6**). For *spc15* and *spc26D* escape mutants, the target sequence is modified through recombination with homologous sequences in the *E. coli* chromosome. Previous studies have demonstrated that Exo-Beta is 100-fold more efficient than the RecABCD pathway in recombining homologous sequences that diverge by 22%

(Martinsohn et al., 2008), a finding that could explain the higher number of this type of recombinant escapers in the presence of the phage repair system. In this case, target mutations are introduced during infection through recombination with an already mutated template DNA. In contrast, the other two types of escape mutations are generated *de novo*. In the case of the target point mutations (*spc9*, *spc40*, *spc45c*, *spc9R*, *spcL4-R* and *spcL6-R* targets), we believe that processing of DSBs exposed after cleavage of the invading DNA by CRISPR nucleases would induce the SOS system (Mo et al., 2021) and lead to the expression of error-prone DNA polymerases (Maslowska et al., 2019). This seems to be particularly important in the case of Cas9 DSB repair, where we found that Pol IV is required for the generation of 99% of the phage escapers (**Fig. 3.3-AB**). However, this polymerase did not affect the mutation pattern (**Fig. 3.4**), suggesting that exact nucleotide changes are related to Exo-Beta activity and/or the degree of Cas9 evasion imparted by a specific mutation, with Pol IV either facilitating or tolerating the incorporation of these changes into the phage genomes. We also found that Exo-Beta can participate in the generation of target deletions via MMEJ during *spc14* targeting, where sequences with 6-12 bp of microhomology are annealed together to excise out the DNA between the microhomology sequences, resulting in deletions of varying sizes (**Table 3.16** and **Table 3.18**). The exact molecular mechanisms behind this mode of DNA repair are not fully understood in *E. coli*. However, since Exo-Beta promotes recombination through the annealing of complementary ssDNA, it is possible that this system could facilitate the annealing of short homologous sequences exposed as ssDNA after Exo

resection at either end of a DSB, resulting in a deletion of the sequence in between the repeats (Kuzminov, 1999; Roy et al., 2020).

A previous study proposed that phage concatemers may be useful for the generation of Cas9 escape mutations in T4 phages (Tao et al., 2018), by facilitating rapid recombination of a cleaved DNA molecule with an intact template. This would increase the number of cleavage-repair rounds and therefore the probability of target mutation. In the absence of phage recombinases, there would be less phage DNA accumulation (due to reduced replication and concatemer formation), fewer possibilities to pair cleaved DNA with an intact repair template and thus a lower target mutation frequency. Although the Red system has been implicated in the generation of λ phage concatemers that facilitate replication and packaging (Enquist and Skalka, 1973), several lines of evidence presented in our work suggest that differences in replication are not responsible for our findings. First, the λ*vir* Δ*exo* mutant phage, which displayed even higher levels of DNA accumulation than wild-type λ*vir* without the addition of any *chi* sites (**Fig. 2.5-A**), was incapable of escaping *spc9* or *spc40* targeting with high frequency (**Fig. 2.4**). Second, the λ*vir* Δ*exo* and λ*vir* Δ*bet* mutant phages exhibited near identical plaquing phenotypes during Cas9 targeting (**Fig. 2.4**), even though the λ*vir* Δ*bet* mutant phage was impaired in DNA replication compared to wild-type phage (**Fig. 2.5-A**), opposite to the λ*vir* Δ*exo* mutant phage that displayed elevated levels of phage DNA accumulation. Third, our qPCR analysis showed that upon addition of a *chi* site in the λ*vir* genome, the wild-type and mutant

phages used in this study all have similar DNA accumulation levels (**Fig. 2.7**). However, at this point, it is not clear to us how the addition of a single *chi* motif affects the replication and recombination of these λ phages. For the Δ*gam* and Δ*red* mutants, it is believed that the added *chi* sequence allows RecABCD recombination to restore concatemer formation (Henderson and Weil, 1975). It is also known that *chi* motifs interfere with the replication and recombination of lambdoid phages that harbor Red-like recombination systems (Henderson and Weil, 1975), an observation that explains the reduction in viral DNA levels for the λ*vir chi*^{1} and λ*vir chi*$^{2\text{-}7}$ phages compared to the wild-type *chi*-less λ*vir* phage (**Fig. 2.14**). Importantly, λ*vir* and λ*vir chi*^{1} phages show identical plaquing efficiencies upon Cas9 targeting (compare **Fig. 2.2** and **Fig. 2.6**), another key observation which confirms that phage DNA levels are not driving the plaquing phenotypes observed in our assays. Nevertheless, due to limitations in our knowledge of λ replication and recombination, we cannot completely exclude a role for concatemer formation by Exo-Beta in the generation of CRISPR escape mutations. The possibility that λ Exo-Beta facilitates the formation of specialized genomic concatemers during replication, that can enhance the speed and efficiency of recombinational repair, is an intriguing hypothesis that requires further study as a follow-up to our ongoing work.

While λ uses the Red recombination system for efficient replication (Enquist and Skalka, 1973), approximately half of the available lambdoid phage genomes do not encode *red* homologues or other dedicated viral recombinases and contain *chi* sites

that allow them to rely instead on the host recombination machinery (RecBCD-RecA) for the formation of concatemers (Bobay et al., 2013). Therefore, Rocha and colleagues have suggested that the presence of dedicated phage recombination systems depends on factors that are unrelated to viral replication (Bobay et al., 2013). On one hand, phage recombinases like Beta are more tolerant to sequence divergence than RecA (Martinsohn et al., 2008), allowing more genetic exchange and mosaicism; on the other, recombination systems take valuable space in the viral genome that could be occupied by more important accessory genes. We believe that our findings reveal an additional role for Red-like recombination systems that favors their preservation within phage genomes: to evade CRISPR immunity and other sequence-specific, DNA-cleaving prokaryotic defense mechanisms (**Fig. 2.17**). Similar to other lambdoid phages that lack the Red system, the introduction of a *chi* site into the λ genome restores efficient viral propagation of Δ*red* phages (Henderson and Weil, 1975). In contrast, neither this site, nor the addition of multiple *chi* sequences or the constitutive recombination activity of RecBC, restores high efficiency of CRISPR escape for phages lacking the Red system. Given that CRISPR-Cas and restriction-modification systems are both prevalent and involve sequence-specific DNA cleavage, we propose that an important driver of the evolution and spread of viral recombinases, which are widely distributed across a large number of both temperate and lytic phage genomes (Lopes et al., 2010), is their ability to repair DNA breaks in a manner that counteracts these defense mechanisms, i.e. through the introduction of escape mutations in the target

sequence. Two recent reports support this idea. One showed that a recombination system related to λ Red present in the IncC conjugative plasmid of *Vibrio cholerae* enables escape from type I CRISPR-Cas cleavage through the introduction of target deletions (Roy et al., 2020). The other report found that the phage T4-encoded UvsX recombinase promotes escape from Cas9 and Cas12a cleavage, also through the generation of deletions (Wu et al., 2021). Importantly, Red systems do not lead to a general surge of mutagenesis, which could be detrimental for both the phage and the host, but rather to a specific increase of mutations at the target site; i.e., only where they matter.

Phages can escape CRISPR immunity by expressing anti-CRISPR proteins (Acrs) that, in most instances, have evolved to specifically interact with Cas nucleases to prevent target cleavage (Stanley and Maxwell, 2018). In contrast to Acrs, λ Red, and possibly other phage-encoded Red-like recombination systems, provide a unique and versatile mechanism that counteracts the activity of CRISPR-Cas (and probably other nucleases involved in anti-phage defense) acting after, instead of before, DNA cleavage, to facilitate phage survival and genetic escape.

7.2 Functional screening of soil metagenomic libraries unearths novel anti-phage defense systems

In Chapter 4, I describe a functional screen that I developed, in collaboration with talented graduate student Christian F. Baca, for the discovery of prokaryotic anti-phage defense systems from environmental DNA (eDNA) (**Fig. 7.5**). Many novel defense systems have been uncovered through bioinformatic exploration of deposited DNA sequences (Doron et al., 2018; Gao et al., 2020; Johnson et al., 2022; Millman et al., 2022). While this approach has been highly effective, success depends on the availability of genomic data. The screening of eDNA libraries does not have this limitation as no previous knowledge of the sequences being surveyed is required for the isolation of defense genes. Related to our approach, one recent study performed a functional screen on genomic libraries of different *E. coli* strains (Vassallo et al., 2022), a method that has the benefit of an almost guaranteed expression of the library inserts. The sequence space screened, however, is limited to the genetic content of this bacterium. The use of eDNA libraries for the isolation of clones with antiviral properties has the caveat that many genes may not be expressed in a heterologous host. On the other hand, eDNA libraries enable exploration of “microbial dark matter” (Rappe and Giovannoni, 2003; Rinke et al., 2013). They provide access to the genetic information of diverse, yet to be discovered, organisms, that do not need to be cultured and that in principle can be isolated from any environment of our planet (Nascimento et al., 2018; Sogin et al., 2006; Thompson et al., 2017). In addition, the discovery of novel defense pathways

can be scaled by screening more than one library, from diverse environmental sources, with a large collection of phages (Maffei et al., 2021). Finally, our screen has the advantage of unearthing not only individual defense systems, but also entire defense islands and/or mobile genetic elements, which may reveal new insights into prokaryotic host-virus conflicts in nature. We envision that this metagenomic screening approach will be applied broadly in both the Marraffini lab and beyond to fuel the discovery of new anti-phage defense systems as well as potential new tools for biotechnology. Work is ongoing in the Marraffini lab on this front, and I eagerly await the discoveries that will be made by the lab's current and next generation of scientists.

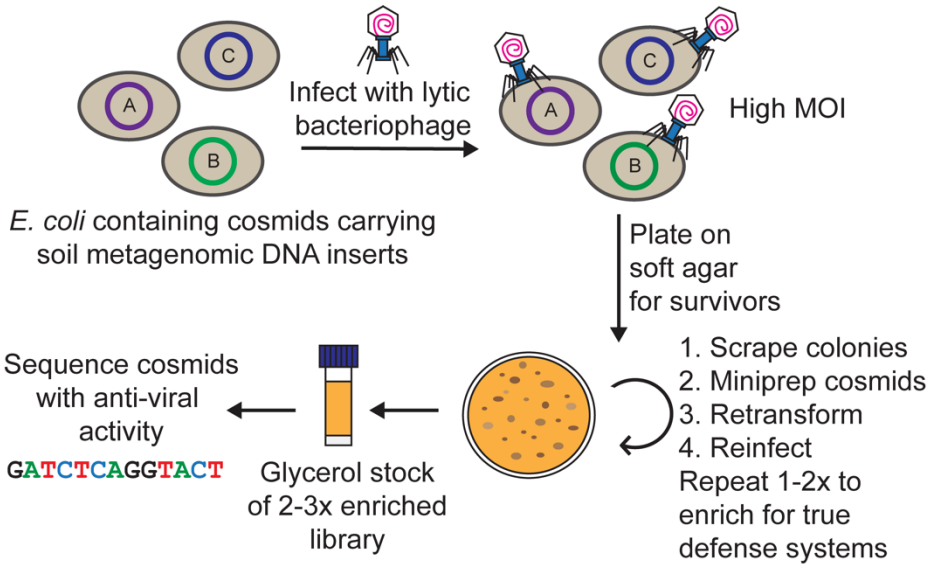


Figure 7.5 Functional selection of soil metagenomic libraries unearths novel anti-phage defense systems

Schematic of the functional selection approach to uncover bacteriophage defense systems in a soil DNA library. First, we subject the host *E. coli* to infection by a lytic bacteriophage at high multiplicity of infection (MOI). Cosmids, carrying library DNA inserts, are extracted from surviving bacterial colonies after each round of infection and retransformed into fresh *E. coli* to generate new libraries, which can then be infected again to enrich for true anti-phage systems carried by library clones.

7.3 The prokaryotic DNA glycosylase Brig1 provides antiviral defense against T-even bacteriophages

Our eDNA library screen, described in Chapter 4, yielded a prokaryotic defense island containing a DNA glycosylase, Brig1 (bacteriophage replication inhibition DNA glycosylase 1), that provides immunity in *E. coli* against T-even bacteriophages by excising alpha-glucosyl-hmC nucleobases in the viral DNA. In Chapters 4-6, I describe a thorough molecular and biochemical characterization of the Brig1 enzyme and report on its homologs in available genetic databases. Our experiments showed that Brig1 did not restrict the propagation of T4 Δ*a-gt* (**Fig. 4.9-A**) nor Bas46-47 coliphages (**Fig. 6.6-B**), whose genomes lack alpha-glucosyl-hmC and instead contain beta-glucosyl-hmC and arabinosyl-hmC nucleobases, respectively. The enzyme also failed to degrade T6 phage DNA, which primarily harbors gentiobiosyl-hmC (**Fig. 6.4-B**). Therefore, it is conceivable that T-even phages have diversified their hmC modification patterns to avoid restriction by DNA glycosylases involved in phage defense such as Brig1. If so, the arms race between phages that modify their DNA and their hosts most likely resulted in the evolution of a larger family of Brig DNA glycosylases with activity against different hmC nucleobases, which probably includes some of the Brig1 homologs that we found associated with other defense genes but failed to provide immunity against T4 and T6 (**Fig. 6.8**). The discovery of new Brig family enzymes is an ongoing future direction in the lab.

While we observed degradation of T4 DNA upon treatment with Brig1 (**Fig. 5.9**), our DNA glycosylase assays with oligonucleotide substrates showed that ssDNA and dsDNA substrates containing a single alpha-glucosyl-hmC nucleobase or dsDNA substrates containing two alpha-glucosyl-hmC nucleobases on opposite strands remained largely uncleaved and full-length even despite overnight incubation with large amounts of Brig1 protein (**Fig. 5.5-A**, **Fig. 5.8-D** and **Fig. 5.8-I**). We therefore propose that Brig1, like members of the uracil DNA glycosylase superfamily (Jacobs and Schar, 2012; Schormann et al., 2014), is primarily a monofunctional DNA glycosylase rather than a bifunctional glycosylase-lyase, which would also nick the DNA phosphate backbone upon base excision (Jacobs and Schar, 2012; Schormann et al., 2014; Zhu, 2009). We believe that Brig1 activity disrupts the lytic cycle of T-even viruses by generating abasic sites throughout the viral genome that: (i) impede phage transcription and/or replication; (ii) lead to spontaneous hydrolysis of the phosphate backbone at the highly reactive abasic sites; and/or (iii) result in DNA interstrand crosslinks and DNA-protein crosslinks due to abasic site reactivity (Thompson and Cortez, 2020). In addition, given that Brig1 can target ssDNA substrates, it would be possible for this enzyme to attack ssDNA intermediates that form during rolling-circle replication of T-even phages. Although we observed low levels of uracil DNA glycosylase activity with Brig1 in vitro (**Fig. 5.5-A**), we attribute this result to the use of high concentrations of the enzyme in our assay. Nonetheless, this weak uracil activity confirms Brig1 as related to the uracil DNA glycosylase

superfamily and as likely having evolved from a not yet identified ancestral uracil DNA glycosylase.

Functionally, but not structurally, Brig1 is related to the restriction enzymes GmrSD and R.PabI, albeit in different ways. Like Brig1, the *E. coli* type IV restriction enzyme GmrSD targets glucosylated hmC-containing DNA in T-even phages (Bair and Black, 2007). Unlike Brig1, however, GmrSD targets both alpha- and beta-glucosyl-hmC. GmrSD contains an HNH endonuclease motif that cleaves phosphodiester bonds to introduce double-strand DNA breaks (Machnicka et al., 2015). Brig1 in contrast possesses a pyrimidine DNA glycosylase pocket with base excision activity on both ssDNA and dsDNA substrates. Interestingly, the T-even phages carry a GmrSD inhibitor, IPI*, which enables viral propagation in the presence of this restriction enzyme in *E. coli* (Bair et al., 2007). In a similar vein, phages with glucosylated cytosines that encounter Brig1 in their native hosts may have evolved inhibitors of Brig1. In support of this prediction, *Bacillus* phages PBS1/PBS2 and Φ29, which incorporate uracil in their genomes and are subject to restriction by host uracil DNA glycosylases involved in base excision repair, encode inhibitors of this enzyme, UGI (Savva and Pearl, 1995; Wang and Mosbaugh, 1989) and p56 (Serrano-Heras et al., 2008; Serrano-Heras et al., 2007), respectively. In contrast to Brig1 and GmrSD, the restriction enzyme R.PabI from the hyperthermophilic archaeon *Pyrococcus abyssi*, targets unmodified DNA and belongs to a family of type II restriction enzymes with a “half-pipe” structural fold (Miyazono et

al., 2014). R.PabI and related enzymes possess adenine DNA glycosylase activity that excises unmodified adenines within a specific DNA sequence: 5'-GATC-3' (Ishikawa et al., 2005; Miyazono et al., 2014). The resultant abasic site ultimately leads to cleavage of the phosphodiester bond, either through heat-promoted beta-elimination (Miyazono et al., 2014), the presence of a secondary AP lyase activity in R.PabI and/or by host AP endonucleases (Fukuyo et al., 2015; Zhang et al., 2017). Therefore, while R.PabI is not a canonical restriction endonuclease, it provides immunity primarily through the generation of a site-specific double-strand DNA break like other type II restriction enzymes. In contrast, Brig1 appears to lack AP lyase activity during base excision and therefore provides anti-phage defense through a different molecular mechanism than both GmrSD and R.PabI. In this sense, Brig1 may be categorized as a restriction DNA glycosylase as opposed to a classical restriction enzyme (**Fig. 7.6**).

One advantage that a glycosylase-based phage-targeting system has over DNA-cleaving restriction endonucleases is that while most classical restriction enzymes exhibit stringent sequence specificity, DNA glycosylases do not require target sequences. This enables Brig1 to cause widespread DNA damage across the entire alpha-glucosylated genome of an infecting T-even phage. Type IV restriction enzymes also exhibit low sequence selectivity, generating double-strand DNA breaks across the phage genome, at or near target nucleobases (Loenen and Raleigh, 2014). Brig1 instead generates abasic sites at target nucleobases, whose highly

reactive chemistry leads to a variety of deleterious DNA lesions. These include spontaneous and endonuclease-assisted ssDNA and dsDNA breaks but also base mismatches and mutation, and replication fork stalling due to DNA-DNA interstrand crosslinks and DNA-protein crosslinks that can be formed at an abasic site to provide a potent roadblock to the progression of DNA/RNA polymerases (Thompson and Cortez, 2020). Furthermore, a double-strand DNA break arising from beta-elimination at an abasic site cannot be repaired as easily as a double-strand break generated by a restriction endonuclease due to a missing base at the site of strand cleavage (**Fig. 7.7**). Finally, Brig1 can attack ssDNA intermediates during viral replication, unlike restriction enzymes that act only on dsDNA.

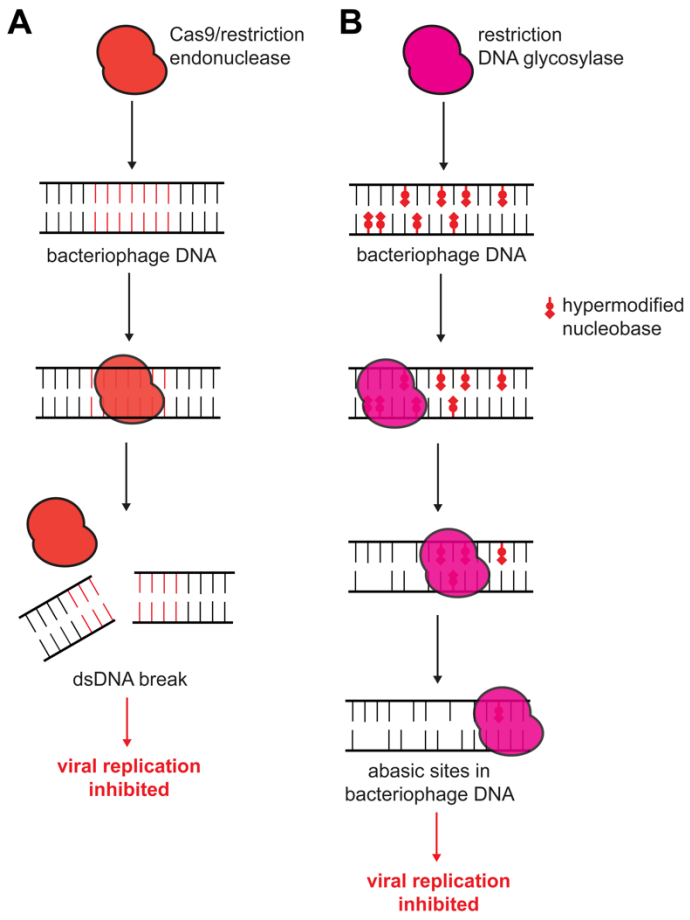


Figure 7.6 Mechanisms of anti-phage defense mediated by restriction enzymes and restriction DNA glycosylases

(**A**) Restriction enzymes and CRISPR-Cas enzymes, e.g., Cas9, are typically endonucleases that recognize a specific target sequence (red) in invading dsDNA. These enzymes cleave the DNA at or near target sequences to generate a dsDNA break. (**B**) Restriction DNA glycosylases, such as Brig1, recognize and excise specific hypermodified nucleobases in both invading ssDNA and dsDNA to generate abasic sites.

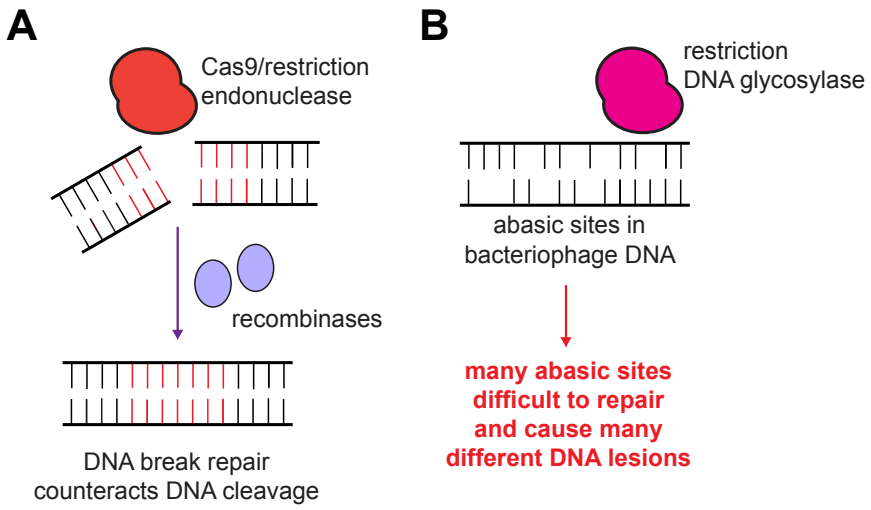


Figure 7.7 Plentiful abasic sites pose a challenge to DNA repair

(**A**) Restriction enzymes and CRISPR-Cas enzymes generate dsDNA breaks that can be repaired by host- and phage-encoded DNA recombinases. (**B**) Abasic sites generated across the phage genome by a restriction DNA glycosylase such as Brig1 may translate to a variety of DNA lesions due to abasic site reactivity, posing a challenge to efficient and faithful DNA repair.

Given that there is no evidence for the misincorporation of alpha-glucosyl-hmC into bacterial genomes in the absence of phage infection, it is unlikely, albeit not impossible, that Brig1 would participate in a host base excision repair pathway dedicated to the removal of these nucleobases. One possibility is that these

nucleobases may arise through the action of DNA-glycosylating interbacterial and phage-encoded toxins, an unlikely scenario since it requires at least a minimum presence of hmC nucleobases in the genome of the host bacterium that a competitor's or invader's alpha-glucosyltransferase enzyme can target (**Fig. 4.9-B**). On the contrary, we believe that Brig1 is a *bona fide* antiviral effector. Supporting this idea, the *a-gt* gene, responsible for the generation of alpha-glucosyl-hmC nucleobases, is widespread across phages infecting diverse hosts (**Table 7.1**), and we found that Brig1 provided immunity against 11/69 phages from the BASEL collection (**Fig. 6.6-A**). Furthermore, many Brig1 homologs are part of phage defense islands, frequently associated with BREX/Pgl genes, but also with toxin-antitoxin cassettes, restriction endonucleases and CRISPR-Cas systems (**Fig. 6.7** and **Fig 6.8**). In these genetic contexts, Brig1 provides an additional layer of immunity against phages containing alpha-glucosyl-hmC that cannot be targeted by other systems present in the phage defense islands, such as CRISPR-Cas and restriction endonucleases and possibly BREX immunity. Although the exact mechanism used by BREX systems to restrict phage infection is unknown, it was recently shown that wild-type T4, but not an *a-gt*/*b-gt* double mutant phage that lacks hmC glucosylation, evades the BREX system present in *E. coli* HS (Gordeeva et al., 2018), suggesting that BREX systems are inhibited by glucosylated nucleobases. In hosts harboring defense islands with restriction enzymes, CRISPR or BREX in addition to Brig1, phages targeted by the latter will not be able to escape through mutations in *a-gt* that eliminate alpha-glucosylation of the viral DNA, since they will now become

susceptible to the defense mechanisms that target non-glucosylated DNA. Brig1 therefore equips the host with a counter strategy to tackle phages that deploy the anti-restriction mechanism of glucosylating their cytosines to evade restriction (**Fig. 7.8**). As mentioned above, the next step of the arms race could involve a change in the phage's nucleobase modification to avoid Brig1 excision and regain virulence.

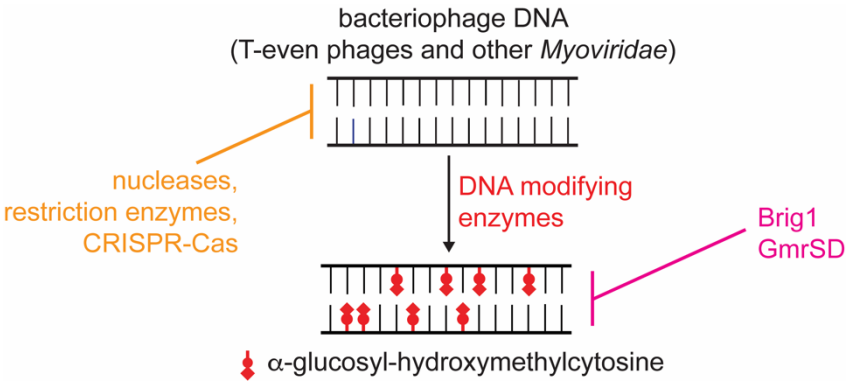


Figure 7.8 Brig1 counters the phage anti-restriction strategy of glucosylating viral nucleobases

T-even phages replace their cytosines with hydroxymethylcytosines, and further modify these nucleobases with glucose adducts. These glucosylated nucleobases prevent recognition and cleavage by host nucleases, restriction enzymes and CRISPR-Cas complexes, thereby enabling viral evasion of host immunity. Brig1, and the type IV restriction enzyme GmrSD, target phages with alpha-glucosyl-hydroxymethylcytosine nucleobases, providing a counter strategy that overcomes the phage's anti-restriction tactic of glucosylating nucleobases.

While alpha-glucosyl-hmC nucleobases are prevalent in T-even phages (**Fig. 6.6-A** and **Table 7.1**), many other different modified nucleobases are present in the genomes of other phages (Hutinet et al., 2019; Khudyakov et al., 1978; Korn et al., 2021; Kropinski et al., 2018; Swinton et al., 1985; Thomas et al., 2018; Weigele and

Raleigh, 2016). These nucleobases could well be the substrates of yet undiscovered DNA glycosylases involved in phage defense such as Brig1. Indeed, I believe new members of the Brig family enzymes, targeting differentially modified hmC nucleobases, are waiting to be discovered in soil bacterial metagenomes. Our study paves the way for exploring this new immunity mechanism with an alternative mode of attacking viral DNA that involves base excision rather than the direct cleavage of sugar-phosphate backbones.

Table 7.1 Protein BLAST of T4 gene *a-gt*, encoding alpha-glucosyltransferase

#	Organism	Query Cover	% identity
1	Escherichia phage T4	100%	100
2	Escherichia phage T4	100%	100
3	Enterobacteria phage T6	100%	99
4	Escherichia phage slur14	100%	99
5	Enterobacteria phage T6	100%	98.75
6	Escherichia phage vB_EcoM_F1	100%	98.75
7	Escherichia coli	100%	98.75
8	Escherichia phage JEP6	100%	98.75
9	Escherichia phage vB_EcoM_FT	100%	98.75
10	Escherichia phage T4	100%	99.75
11	Tequatrovirus T4	100%	99.5
12	Escherichia phage T4	100%	99.5
13	Yersinia phage fPS-90	100%	98.5
14	Escherichia phage HY03	100%	98.25
15	Escherichia phage T4	98%	99.75
16	Escherichia phage vB_EcoM_ACG-C40	100%	93.75
17	Escherichia phage vB_EcoM_G8	100%	91.75
18	Tequatrovirus RB14	100%	91.5
19	Shigella flexneri	100%	91.75
20	Escherichia phage PE37	100%	91.5
21	Enterobacteriaceae	100%	91.25
22	Enterobacteria phage RB27	100%	91.25

23	Escherichia phage vB_EcoM_G50	100%	91.25
24	Escherichia phage teqsoen	100%	91
25	Escherichia phage UoN_LDJ77_1	100%	91.5
26	Salmonella phage GRNsp7	100%	91.25
27	Shigella sonnei	100%	91.25
28	Escherichia phage AugustSocin	100%	91.25
29	Escherichia phage ECO4	100%	91.25
30	Escherichia phage slur02	100%	91
31	Enterobacteria phage RB51	100%	91.25
32	Salmonella enterica subsp. enterica serovar Braenderup	100%	91.25
33	Escherichia phage vB_EcoM_KAW1E185	100%	91.25
34	Escherichia phage RB3	100%	91.25
35	Escherichia phage vB_EcoM_Shinka	100%	91.25
36	Escherichia phage vB_EcoM_DalCa	100%	91.25
37	Escherichia phage UFV-AREG1	100%	91.25
38	Citrobacter phage PhiZZ6	100%	91.25
39	Shigella phage ESH35	100%	91.25
40	Escherichia phage vB_EcoM-S1P5QW	100%	91.25
41	Escherichia coli	100%	91
42	Escherichia phage vB_EcoM_G4498	100%	91
43	Escherichia phage UTI-E4	100%	91
44	Shigella phage KRT47	100%	91
45	Shigella phage ESh29	100%	91.25
46	Escherichia phage PP01	100%	91.25
47	Escherichia phage vB_EcoM_Kelasse	100%	91.25
48	Escherichia phage KarlGJung	100%	91
49	Escherichia phage teqhal	100%	91
50	Shigella phage ESh17	100%	91
51	Escherichia coli	100%	91
52	Shigella flexneri	100%	91
53	Shigella phage Shfl2	100%	91
54	Shigella phage Sf23	100%	91
55	Yersinia phage PYps14T	100%	91.25
56	Shigella flexneri	100%	90.75
57	Escherichia phage vB_EcoM-Sa45lw	100%	91
58	Escherichia phage T2	100%	91
59	Bacillus cereus	100%	90.75
60	Escherichia coli	100%	90.75
61	Shigella flexneri	100%	90.75

62	Shigella sonnei	100%	91
63	Shigella sonnei	100%	91.25
64	Escherichia phage TadeuszReichstein	100%	91.25
65	Escherichia phage vB_EcoM_SP1	100%	90.75
66	Escherichia phage NiC89	100%	90.75
67	Escherichia coli	100%	91
68	Escherichia phage vB_vPM_PD112	100%	91
69	Escherichia phage vB_EcoM-101117BS1	100%	90.75
70	Phage NBEco003	100%	90.75
71	Escherichia coli	100%	91
72	Yersinia phage PYps55T	100%	90.75
73	Yersinia phage fPS-2	100%	91
74	Escherichia phage vB_EcoM_ESCO58	100%	91
75	Escherichia phage Paracelsus	100%	91
76	Escherichia phage vB_EcoM_G2540-3	100%	91
77	Shigella phage Sf24	100%	90.75
78	Escherichia phage vB_EcoM-RPN226	100%	91
79	Escherichia phage JLBYU24	100%	90.75
80	Shigella phage Sf21	100%	91
81	Shigella phage KNP5	100%	90.5
82	Shigella phage ESh16	100%	90.5
83	Escherichia phage vB_EcoM_Lutter	100%	90.75
84	Shigella phage CT01	100%	90.75
85	Escherichia phage vB_EcoM_Nami	100%	90.75
86	Escherichia coli	100%	90.5
87	Escherichia phage vB_EcoM_G10400	100%	90.75
88	Enterobacteria phage vB_EcoM_IME339	100%	90.5
89	Escherichia coli	100%	90.75
90	Escherichia phage EcNP1	100%	90.5
91	Yersinia phage PYps32T	100%	90.75
92	Escherichia phage UGKSEcP2	100%	90.25
93	Escherichia coli	98%	91.88
94	Escherichia coli	98%	91.62
95	Escherichia phage W143	98%	91.37
96	Escherichia phage N2	98%	91.12
97	Salmonella phage Lv5cm	98%	90.86
98	Escherichia phage vB_EcoM-BECP11	91%	89.86
99	Escherichia phage AR1	88%	89.58
100	Escherichia coli	87%	91.12

101	Escherichia coli	82%	91.84
102	Salmonella enterica subsp. enterica serovar Braenderup	78%	91.4
103	Escherichia coli	77%	91.56
104	Escherichia coli	75%	91.69
105	Klebsiella phage vB_KpM-Wobble	100%	66.08
106	Klebsiella phage JIPh_Kp122	100%	66.08
107	Klebsiella phage KMI13	100%	66.08
108	Klebsiella phage PKO111	100%	66.08
109	Klebsiella phage KP1	100%	66.08
110	Klebsiella phage vB_KaeM_Nispero	100%	66.08
111	Klebsiella phage pKp20	100%	66.08
112	Klebsiella phage Mineola	100%	66.08
113	Kelbsiella phage vB_KaeM_Boboto	100%	66.08
114	Salmonella phage PSE-D1	100%	66.08
115	Klebsiella variicola	100%	65.84
116	Klebsiella phage KP179	100%	65.84
117	Klebsiella phage JD18	100%	65.84
118	Klebsiella phage vB_KaeM_Merci	100%	65.84
119	Klebsiella phage vB_KaeM_LilPanda	100%	65.59
120	Klebsiella phage KPV15	100%	65.59
121	Klebsiella phage KP13MC5-5	99%	65.75
122	Klebsiella phage UTI-K1	100%	65.59
123	Klebsiella phage vB_KpnM_BovinicusUrsus	100%	65.59
124	Klebsiella phage vB_KpnM_17-11	100%	65.59
125	Klebsiella phage vB_KpnM_KpV477	100%	65.59
126	Klebsiella phage vB_KaeM_Shinkou	100%	65.59
127	Klebsiella phage KpnM6E1	100%	65.34
128	Klebsiella phage KP182	100%	65.59
129	Klebsiella pneumoniae	100%	65.59
130	Bacteriophage sp.	100%	65.34
131	Klebsiella phage Metamorpho	100%	64
132	Bacteriophage sp.	95%	65.35
133	Erwinia phage Cronus	100%	56.47
134	Bacteriophage sp.	46%	64.13
135	Bradyrhizobium sp. Ce-3	27%	98.18
136	Maribacter sp.	100%	32.28
137	uncultured Caudovirales phage	99%	31.28
138	Verrucomicrobiales bacterium	99%	28.57
139	Xanthomarina gelatinilytica	59%	34.55

140	Escherichia coli	19%	89.61
141	Podoviridae sp.	99%	27.92
142	Gammaproteobacteria bacterium TMED36	99%	29.11
143	Dyella japonica	18%	93.06
144	Salmonella enterica subsp. enterica serovar Braenderup	17%	88.41
145	Candidatus Endolissoclinum sp. TMED37	98%	24.94
146	Escherichia coli	14%	89.29
147	Prokaryotic dsDNA virus sp.	100%	28.43
148	Xanthomarina gelatinilytica	28%	40.83
149	Bacillus phage G	99%	24.27
150	Cyanophage S-TIM54	87%	25.39
151	Bacillus phage G	98%	21.84
152	Cyanophage S-TIM5	87%	23.44
153	Cyanophage S-TIM66	87%	23.44

7.4 ADP-ribosylglycohydrolases in anti-phage defense or Brig1 regulation

In Chapter 4, the Brig1 operon we isolated does not harbor any of the defense systems mentioned above (**Fig. 4.3-C**). Instead, Brig1 is flanked by a protein with predicted ADP-ribosyl glycohydrolase activity. Interestingly, these enzymes remove ADP-ribose groups from proteins and DNA to counter ADP ribosyltransferase (ART) toxins, commonly involved in bacteria-phage conflicts (Aravind et al., 2015). Phage T4 harbors three ARTs that ADP-ribosylate host proteins to facilitate the viral lytic cycle (Alawneh et al., 2016; Tiemann et al., 2004). Although not required for immunity in our assays, it is possible that in other conditions, or after infection by a different phage expressing an ART, the ADP-ribosyl glycohydrolase adjacent to Brig1 is required to de-toxify the host from ADP-ribosylation and facilitate recovery after

infection has been cleared by Brig1 or another anti-phage system. A phage-encoded ART may even serve to inhibit Brig1 activity through ADP-ribosylation of target amino acids in Brig1. In this case, the ADP-ribosyl glycohydrolase can enable Brig1 to regain its anti-phage activity through removal of the inactivating ADP-ribose modifications, explaining its co-occurrence with Brig1 in multiple genetic loci (**Fig. 6.7**).

The ADP-ribosyl glycosylhydrolase-encoding protein could also regulate Brig1 activity in other ways. Indeed, the C-terminal domain of this protein is a dual-specificity protein phosphatase that can dephosphorylate phospho-serine/threonine and phosphotyrosine residues (Pulido and Lang, 2019). This domain may counter phosphorylation of Brig1 and/or other host proteins by host- and phage-encoded protein kinases or regulate its own adjacent ADP-ribosyl glycohydrolase activity. Another intriguing possibility is that the ADP-ribosyl glycohydrolase participates in phage defense akin to Brig1, i.e., by inhibiting the replication of phages carrying ADP-ribosylated nucleobases. Future work in the lab will focus on the biological roles of ADP-ribsoyl glycohydrolases that colocalize with Brig1.

CHAPTER 8. OUTLOOK

In this thesis, I present two cases where DNA repair mechanisms have been co-opted to execute non-canonical immune and counter-immune roles in prokaryotic host-virus genetic conflicts. I speculate that there remain many more such instances waiting to be discovered, especially given the multitude of bacterial immune systems that target the DNA of invading phages and plasmids (Georjon and Bernheim, 2023), as described in Chapter 1. Since the targeted cleavage of foreign DNA is a widespread mechanism of anti-phage immunity (Georjon and Bernheim, 2023), executed predominantly by RM and CRISPR-Cas systems (Tesson et al., 2022) (**Fig. 1.6**), it is quite possible that phage recombination systems, such as λ Red, are specifically adapted to countering DNA cleavage by RM and CRISPR, and enabling phage survival in the face of targeted DNA degradation. Indeed, work presented in this thesis suggests that λ Red is more effective than the host's RecBCD-RecA recombination pathway in promoting CRISPR-Cas evasion and generating large numbers of escape phages (**Fig. 2.15**), even when the bacterial system is rigged to undergo constitutive recombination at DNA breaks (**Fig. 2.16**). Many phage recombination systems such as the Red-like systems promote strand annealing (Lopes et al., 2010; Mosberg et al., 2010) over a RecA-like strand invasion mechanism during homology-mediated target search. Future work may focus broadly on whether strand annealing mechanisms yield faster and more efficient break repair and are therefore preferred by phages not only to facilitate fast viral lifecycles but also to promote rapid recombinational repair when encountering the myriad anti-

phage defenses that cleave viral DNA. Such studies would need to generalize our findings to Red-like and other phage recombination systems beyond λ Red. One possibility is to perform single-molecule experiments (Sternberg et al., 2014) that assay the kinetics of in vitro DNA break repair using different purified and in vitro reconstituted recombination systems, to probe what happens after phage DNA is cleaved by a restriction endonuclease or by Cas9. Another interesting hypothesis that prompts further exploration is whether λ Red and similar systems promote the formation of a complex network of phage genome concatemers that rapidly pairs homologous templates to facilitate recombinational repair. Experiments to assay whether concatemers drive repair of CRISPR-cleaved phage DNA are difficult but may include electron microscopy or cryo-electron tomography of CRISPR-targeting and non-targeting cells infected with phage, with and without its encoded recombination system, to visualize viral concatemers, as has been done previously to visualize replication intermediates formed during the replicative cycle of phage ΦX (Koths and Dressler, 1978). One can also use chromosome conformation capture techniques such as Hi-C (Le et al., 2013) to determine genome-wide points of contact during concatemer formation and ask whether these promote break repair and are dependent on phage recombination.

One of the speculations I put forward in this thesis is that CRISPR evasion by λ Red may be a key driver for the selection of Red-like and other phage-encoded recombination systems in bacteriophage genomes. Indeed, the competition

experiment between Red+ and Red- phages found a strong selection for Red+ phages in the presence of Cas9 targeting (**Fig. 2.17**). In the absence of CRISPR, neither phage was selected, a result that supports the hypothesis that Red-like systems are evolutionarily selected specifically for their roles in immune evasion. A recent study showed that the UvsX recombinase encoded by phage T4 also enables evasion of Cas9 targeting, by driving minihomology-mediated recombination to eliminate the Cas9 target sequence (Wu et al., 2021). Another study found that Red-like systems, encoding genes homologous to λ Exo and Beta, in IncC and SXT genetic elements of *Vibrio cholerae* mediated similar target deletions to drive evasion of the host's type I-F CRISPR-Cas system (Roy et al., 2020). Together, these provide increasing evidence that phage-encoded recombination systems possess a major role in protecting viral genomes against targeted cleavage by bacterial defense systems. A future study may involve bioinformatics to examine whether phages that are more likely to encounter RM or CRISPR-Cas in their native host strains are also more likely to encode their own recombination systems as opposed to relying on the host's system for DNA repair. Such studies, however, may be compounded by the prevalence of RM systems across prokaryotic genomes, the presence of unidentified Acrs and RM inhibitors in phage genomes, and by the difficulty of accurately predicting host-phage pairs based on sequences alone.

An interesting caveat of the competition experiment is that even though phages were infected at a high initial multiplicity of infection (MOI 10), we still see a

complete and total selection of Red+ phages (**Fig. 2.17**). This is counterintuitive to the idea that Red proteins are available as "public goods" in cells infected with Red+ phages, where Red- phages can use Red proteins if they infect the same cell. Here, the expectation is that at high MOI we would see a small selection, if any, for Red+ phages as both Red+ and Red- phages can access the beneficial Red proteins as public goods. What we observed was contrary to this expectation (**Fig. 2.17**), and one possibility could be that there are unknown superinfection exclusion mechanisms at play during λ infection that limit infection by multiple phages even at high culture MOIs. Another hypothesis, perhaps more interesting and not mutually exclusive, is that infecting phage genomes form infection foci within the cell, with localized production of proteins that act more efficiently within proximity to their corresponding phage genomes within an infection locus. Super-resolution microscopy of cells infected with fluorescently labeled phage genomes, using fluorescent *tetO*-TetR operator-reporter systems (Williams et al., 2023), and labeled phage proteins would be necessary to confirm or exclude this hypothesis.

In our study, we found that λ Red promotes CRISPR-Cas escape mutations, including point mutations, deletions, and host recombinants (**Fig. 2.2**). The origin of these mutations is not always understood, although our results show that the error-prone DNA polymerase Pol IV may play a role in the introduction of target point mutations (**Fig 3.3**). However, since Pol IV does not affect the nature of the observed mutations (**Fig. 3.4**), it remains unclear what processes drive the generation of these

mutations in the first place, while Pol IV tolerates the mutations during fill-in synthesis. The deletion mutations observed are all mediated by 6-12 bp microhomology sequences (**Table 3.16** and **Table 3.18**), and any additional factors that may promote microhomology-mediated end joining (MMEJ) in *E. coli* are currently poorly defined (Chayot et al., 2010). To identify host factors that play a role in Red-mediated CRISPR evasion, future work may employ an unbiased transposon sequencing (Tn-seq) library screen with Cas9 targeting of phage λ using *spc14*, where the predominant mode of phage escape is through target deletions generated by MMEJ (**Table 3.16** and **Table 3.18**). A Tn-seq library of phage mutants can also be used to identify any phage factors that mediate CRISPR-Cas evasion in addition to or in synergy with λ Red. The phage Tn-seq library would first need to be generated in a CRISPR-, RM- "defenseless" strain of *E. coli* (Maffei et al., 2021), which would enable the knockout of both nonessential and any anti-defense phage proteins in the λ genome. This Tn-seq phage library can be generated in *red*+ and *red*- genetic backgrounds and then challenged with Cas9 targeting using different spacers. Broad-based approaches such as these may provide further examples of DNA repair proteins, encoded by host or phage, being co-opted by the virus to counteract CRISPR-Cas targeting. Alternatively, other phage or host recombination proteins, such as RecJQ (Yasmin et al., 2021) or SbcCD (Connelly et al., 1998), can be knocked out or knocked down in a targeted manner and assayed for their role in the promotion or evasion of CRISPR-Cas immunity.

We found that while λ Red mediates evasion of Cas9 targeting for most spacers that we tested (so called “weak” spacers”), there are cases of strong targeting, such as with *spc45*, where λ Red is unable to provide meaningful protection against Cas9 cleavage (**Fig. 2.9** and **Fig. 3.7**). This raises the possibility of cooperation between Red-like recombination systems, which act after the CRISPR-Cas immune response has cleaved its target DNA, and phage-encoded Acr proteins, which directly bind to Cas9 and other Cas effectors to prevent target binding or cleavage (Davidson et al., 2020) (**Fig. 1.13**). Studies of phage-encoded Acrs have shown that at low doses these proteins partially inhibit CRISPR-Cas activity (Borges et al., 2018; Landsberger et al., 2018), effectively rendering a “strong” spacer “weak” but without fully eliminating cleavage and immunity (**Fig. 1.14-AB**). Weakened Cas cleavage via Acrs may provide a route for phage escape through λ Red-mediated repair of the CRISPR-cleaved DNA, through recombination with homologous phage DNA templates that are yet to be cleaved due to Acr-mediated inhibition. This synergy between λ Red and Acrs, and even other anti-restriction mechanisms such as DNA modifications (**Fig. 1.12-C**), is an exciting avenue for future study – akin to synergy between CRISPR-Cas and immune systems such as RM (Maguin et al., 2022) – and would shed light on the complex nature and full extent of phages’ anti-CRISPR programs.

Along the same lines, but from the bacterial perspective of the arms race, an intriguing follow-up is to test for synergy between CRISPR-Cas and retron systems

that are activated by RecBCD inhibition (**Fig. 1.5-C**). The Ec48 and Se72 retron systems both guard RecBCD and are activated by λ Gam-mediated inhibition of RecBCD (Millman et al., 2020a; Stokar-Avihail et al., 2023) (**Fig. 1.5-C**). Phages can evade these retron systems by deleting Gam (Millman et al., 2020a; Stokar-Avihail et al., 2023) (**Fig. 1.12-A**). From the results presented in this thesis, we know that deletion of Gam sensitizes phage λ to CRISPR-Cas targeting and reduces the frequency of CRISPR escape mutants (**Fig. 2.12**). Therefore, it follows that CRISPR-Cas and retron-mediated immunity may synergize, wherein, by exploiting opposing vulnerabilities, each system prevents the rise of escape phages of the other system. A well-designed study would couple mechanistic and molecular details of this interaction with a bioinformatic survey of the co-occurrence of DNA-targeting type I and II CRISPR-Cas systems and homologs of RecBCD-guarding Ec48- and Se72-family retrons across prokaryotic genomes.

Finally, our study on λ Red only looks at the role of DNA repair in countering CRISPR interference but ignores the acquisition phase of CRISPR-Cas immunity (**Fig. 1.9**), where new spacers are acquired against an invading phage or plasmid to provide adaptive immunity (Marraffini, 2015). How the presence or absence of λ Red affects the acquisition of spacers against phage λ is an important next question that stems from this work, to gain a fuller picture of how Red-like recombination systems affect the overall CRISPR-Cas immune response in bacteria. Previous studies have shown that RecBCD drives CRISPR spacer acquisition through DNA degradation

and the generation of free DNA ends, which favors the acquisition of new spacers from foreign DNA that lacks *chi* sites (Levy et al., 2015; Modell et al., 2017). Through Gam-mediated inhibition of RecBCD and Exo-Beta driven recombination, λ Red may impact rates and patterns of spacer acquisition from both host and viral DNA. One challenge in addressing this question in *E. coli* is that "naïve" spacer acquisition (the acquisition of new spacers without a preexisting targeting spacer) has not been observed in the laboratory for neither the type II-A nor the type I-E CRISPR-Cas system I worked with (Strotskaya et al., 2017). Nonetheless, "primed" spacer acquisition, where a targeting spacer leads to the acquisition of one or more additional targeting spacers (Nussenzweig et al., 2019; Strotskaya et al., 2017), may serve as a substitute to study the impact, if any, of λ Red or any other phage-encoded recombination system on the rate and pattern of CRISPR spacer acquisition.

In the second half of my thesis, I describe the discovery of Brig1, a bacterial DNA glycosylase that provides defense against T-even phages. Brig1 exemplifies DNA repair repurposing by bacteria instead of phages, suggesting that co-option of DNA repair proteins in immune contexts is more widespread in the prokaryotic domain than previously appreciated. Since most antiviral systems that target foreign DNA encode endonucleases that cleave phage DNA (Georjon and Bernheim, 2023) (**Fig. 1.8**), the discovery of base excision modules in prokaryotic immunity unveils a new mechanism for attacking foreign DNA that relies solely on the removal of

modified nucleobases from viral DNA rather than the cleavage of DNA backbones. Our preliminary analysis of Brig1 homologs yielded only a small group of proteins (**Fig. 6.7**), but many of these are encoded in bacterial defense islands or part of BREX defense systems (**Fig. 6.8** and **Fig. 6.9**). Future studies will explore the role of Brig1 in these defense islands, especially focusing on any synergy between Brig1 and defense systems such as BREX and CRISPR-Cas which are typically disabled by DNA hypermodifications (Bryson et al., 2015; Gordeeva et al., 2018; Liu et al., 2020b; Vlot et al., 2018). Previous work from the Marraffini lab has already established a synergy between RM and CRISPR-Cas systems, wherein DNA cleavage by restriction endonucleases promotes CRISPR spacer acquisition (Maguin et al., 2022). Building on this, it would be interesting to examine whether Brig1 may promote CRISPR spacer acquisition and/or counter the rise of CRISPR escape phages.

Although we found only 42 non-redundant homologs of Brig1 (**Fig. 6.7**), denoted as “hypothetical proteins” on NCBI, suggesting that Brig1 itself may be rare, we cannot exclude the possibility that a much larger collection of homologs is present among unsequenced bacteria in the soil. Indeed, since sequence homology searches are often quite limiting, there are likely several other Brig family DNA glycosylases labeled as “hypothetical proteins” and hiding in plain sight in online genetic databases. While sequence similarity only yields a limited collection of Brig1 proteins, structural homology using HHPred (Zimmermann et al., 2018) and AlphaFold2

(Jumper et al., 2021) suggest that Brig1 is related to the superfamily of uracil DNA glycosylases. To discover the true breadth of Brig1 structural homologs, we would need to develop a structure-guided discovery algorithm that compares the structure of Brig1 to a large collection of predicted structures of unknown hypothetical proteins found within gene neighborhoods of all bacterial defense islands extracted from online genetic databases. Such an ambitious approach is currently too computationally intensive to perform, although the landscape for undertaking such an endeavor may look feasible a few years from now. For now, a more modest approach may entail a broad sequence homology-based search for DNA glycosylases in bacterial defense islands and cloning them, along with their entire defense operons, into *E. coli* to test for any heterologous anti-phage activity. Such an approach can also focus more broadly on DNA repair modules across bacterial defense islands as well as in phage anti-immune loci (such as *acr* loci) to delve deeper into the co-option and roles of DNA repair proteins in the bacteria-phage arms race. An interesting possibility is the co-option of DNA glycosylase domains as toxic effector modules in abortive infection systems, such as CBASS, Pycsar, Avs or retrons, wherein glycosylase domains mediate excision of unmodified, canonical DNA nucleobases to indiscriminately damage both host and viral DNA.

Perhaps the simplest and most immediate approach to search for Brig1 homologs would be to repeat the screen using soil metagenomic libraries from nearby regions or other parts of Arizona and the United States. Such libraries are

already available in Dr. Sean Brady's laboratory at the Rockefeller University. Given the relative ease with which Brig1 was isolated during one of our first attempts at library screening, this approach may indeed yield more homologs. Libraries can also be screened using PCR with degenerate primers to identify and subsequently fish out any library clones that carry homologs. Furthermore, the library screen can be performed with phages carrying different DNA modifications, such as T4 Δ*b-gt* which carries only beta-glucosylated hmC nucleobases, T6 which carries gentiobiosyl-hmC nucleobases (Kuno and Lehman, 1962; Lehman and Pratt, 1960), Bas46-47 which carry arabinosyl-hmC nucleobases (Maffei et al., 2021) and BASEL phages that carry deazaguanosine nucleobases (Maffei et al., 2021). Screening with phages carrying DNA modifications may not only uncover new DNA glycosylases but also novel type IV RM systems and other bacterial defense systems that target modified DNA. Another possibility is the discovery of abortive infection systems that sense modified viral nucleobases as PAMPs. The most pressing next step on the screening front is to do a library screen with the T4 mutant phage, T4 Δ*a-gt* Δ*b-gt*. This T4 mutant lacks alpha- and beta-glucosyltransferases and therefore carries only hmC nucleobases (Lehman and Pratt, 1960). T-even phages always glycosylate hmC nucleobases as an anti-restriction mechanism (Bryson et al., 2015; Vlot et al., 2018), suggesting that defense systems, such as type IV RM systems, that target hmC-containing DNA are plentiful among bacteria. Screening metagenomic libraries with T4 Δ*a-gt* Δ*b-gt* is therefore likely to uncover many new bacterial defense systems that target hmC. This

screening project is currently being undertaken by an up-and-coming talented graduate student in the lab: Adriana Mejia.

A major aim in discovering new bacterial immune systems is their potential usefulness as tools for molecular biology and biotechnology. On this front, Brig1 holds great promise. In eukaryotes, hmC is a common DNA mark in neurons and embryonic stem cells that can regulate transcription (Ficz et al., 2011; Mellen et al., 2012). We are exploring the use of Brig1 as a tool for mapping these DNA modifications in eukaryotic cells via next-generation sequencing. Such a technology would first require extraction of DNA from cells and then treating the DNA with purified T4 alpha-glucosyltransferase to convert hmC to alpha-glucosyl-hmC. This will be followed by Brig1 treatment to generate abasic sites. An already developed technology for the next-generation sequencing of abasic sites can then be employed to map hmC modifications across the genome (Liu et al., 2019b). Future work in the Marraffini lab will focus on the development of this sequencing technology.

On a final note, the metagenomic screening pipeline that I developed promises to be a major discovery tool for the Marraffini lab and I am excited to see what new projects stem from applying or modifying this approach. I hope that future work at this frontier expands our current understanding of bacterial immune mechanisms and continues to uncover exciting new roles for DNA repair proteins in prokaryotic host-virus conflicts.

CHAPTER 9. MATERIALS AND METHODS

9.1 Bacterial strains and growth conditions

Cultivation of *E. coli* K-12 MG1655 (Guyer et al., 1981), *E. coli* EC100 (Lucigen), *E.coli* K-12 BW25113 (Datsenko and Wanner, 2000) and all other related strains were carried out in lysogeny broth (LB) media (BD Difco LB Broth, Lennox, BD 240220 for LB liquid media; BD Difco LB Agar, Miller, BD 244510 for LB agar plates) at 37°C (30°C in certain cases) and for liquid cultures with shaking. Overnight cultures were inoculated from single bacterial colonies. Wherever applicable, media were supplemented with chloramphenicol (GoldBio, C-105-100) at 12.5 μg/ml (for cosmids) or 25 μg/ml (for plasmids), spectinomycin (GoldBio S-140-50) at 50 μg/ml, kanamycin (GoldBio, K-120-100) at 50 μg/ml, carbenicillin (GoldBio, C-103-25) at 100 μg/mL and/or tetracycline (Sigma-Aldrich, T7660-25G) at 5 μg/mL to ensure cosmid, plasmid or deletion strain maintenance. *E. coli* K-12 BW25113 Keio knockout strains were obtained from the Coli Genetic Stock Center at Yale University (Baba et al., 2006). The type I-E CRISPR interference strain *E. coli* K-12 MG1655 ACT-01 was a generous gift from Chris A. Voigt at MIT (Caliando and Voigt, 2015). The type I-E CRISPR-Cas strain *E. coli* KD263 and KD263 λ-targeting strains (λ_*E4-R*, λ_*L1-R*, λ_*L4-R*, λ_*L6-R*) were generous gifts from Konstantin Severinov and Ekaterina Semenova at Rutgers University (Strotskaya et al., 2017). The identity of the spacer within the KD263 CRISPR array was confirmed by amplification with primers JW3100 and JW3101 (see **Table 9.1**) followed by Sanger sequencing. *E. coli* K-12 MG1655

Δ9 (a strain with the 9 *E. coli* prophages deleted from the MG1655 genome) was a generous gift from Thomas K. Wood at the Pennsylvania State University (Wang et al., 2010). The *E. coli* K-12 MG1655::λ lysogen was obtained from J. W. Roberts. Miniprepped plasmids (prepared by QIAprep Spin Miniprep Kit, QIAGEN, 27104) were cloned into chemically competent *E. coli* EC100 cells, electrocompetent *E. coli* EC100 cells or rubidium chloride chemically competent *E. coli* K-12 MG1655 cells. For *E. coli* K-12 BW25113 and Keio knockout strains and strains with two plasmid combinations, existing strains were first made electrocompetent and then transformed with cosmid or plasmid through electroporation (1 mm Bio-Rad Gene Pulser cuvette at 1.8 kV). The bacterial strains used in this study are listed in **Table 9.2**.

9.2 Plasmid construction

Cloning was performed with chemically competent *E. coli* EC100 cells, electrocompetent *E. coli* EC100 cells or rubidium chloride chemically competent *E. coli* K-12 MG1655 cells. For chemical transformation, 20-200 ng of plasmid DNA (prepared by QIAprep Spin Miniprep Kit, QIAGEN, 27104) was mixed with 50 μL of competent cells and incubated on ice for 20-30 minutes. Cells were heat shocked in a 42°C water bath for 45 seconds and then placed back on ice for 3 minutes. Cells were then resuspended in 300 μL room temperature LB medium and incubated at 30°C or 37°C (depending on plasmid) for 1-2 hours. After the outgrowth period, the entire volume of transformed cells was spread on an LB agar plate with the

appropriate antibiotic(s) for plasmid selection and incubated at 30°C or 37°C overnight. The next day, individual colonies were picked and grown in 3 mL LB medium with appropriate antibiotic(s) for plasmid maintenance. To store plasmid strains, 900 μL of overnight culture was mixed with 100 μL dimethyl sulfoxide (Sigma-Aldrich, D2650-100ML) and frozen at -80°C. These frozen stocks were streaked on LB agar plates with appropriate antibiotic(s) to single colonies for use in experiments. For electrocompetent *E. coli* EC100. *E. coli* K-12 BW25113 and Keio knockout strains, the ACT-01 strain and *E. coli* strains with two plasmid combinations that were difficult to transform simultaneously, existing strains were first made electrocompetent when necessary and then transformed with plasmid. To make cells electrocompetent, 1.4 mL of 3 mL overnight stationary cultures were spun down in 1.5 mL Eppendorf tubes at 10,000 rpm for 1 minute in a tabletop microcentrifuge at 4°C. After discarding the supernatant, cells were washed 2 times with sterile cold water. To wash cells, pelleted cells were resuspended in 1 mL sterile cold water, spun down at 10,000 rpm for 1 minute in a tabletop microcentrifuge at 4°C and supernatant discarded. Cells were resuspended in 150 μL cold 10% glycerol (Fisher Scientific, G33-500) for use in electroporation. For electro-transformation, 50 μL of cells were mixed with 50–150 ng of plasmid DNA (prepared by QIAprep Spin Miniprep Kit, QIAGEN, 27104). Cells were electroporated using a 0.1 cm gap Gene Pulser electroporation cuvette (Bio-Rad, 165-2089) at 1.8 kV in a Bio-Rad Gene Pulser Xcell system. Electroporated cells were immediately resuspended in 300 μL of room temperature LB medium. Cells were recovered at 30°C or 37°C for 1-2 hours

before being plated on LB agar (BD Difco LB Agar, Miller, BD 244510) with appropriate antibiotic(s) and incubated at 30°C or 37°C overnight. The plasmids and cosmids used in this thesis are listed in **Table 9.3**.

9.3 Gibson assembly

For Gibson assemblies (Gibson et al., 2009), 25-100 ng of the largest dsDNA fragment was combined with equimolar volumes of the smaller fragment(s) in a total volume of 5 μL in nuclease-free water. Reaction mixtures were prepared on ice and mixed with 15 μL of Gibson assembly master mix, pipette mixed and incubated at 50°C for 1 hour in a thermal cycler. Gibson reactions were transformed into chemically competent *E. coli* EC100 cells (Lucigen) or rubidium chloride chemically competent *E. coli* K-12 MG1655 cells by mixing 5 μL Gibson reaction with 50 μL cells and following the transformation protocol for chemically competent cells outlined in the section above.

9.4 Oligo cloning

Oligo cloning was used to create a repeat-spacer-repeat CRISPR array with a desired spacer following a protocol previously described by the Marraffini lab (Jiang et al., 2013). Briefly, we used a BsaI restriction digest cloning approach. Parent type II-A CRISPR array-containing plasmids with a repeat-spacer-repeat carried a 30 bp spacer sequence with two BsaI cut sites at either end. To set up the BsaI plasmid

digest, we mixed 42 μL of the parent CRISPR plasmid (40-60 ng/μL) with 6 μL BsaI-HF (NEB, R3535L), 6 μL NEB CutSmart buffer and 6 μL nuclease-free water. The restriction digest reaction was incubated at 37°C for approximately 6 hours. Two IDT oligonucleotides comprised the type II-A CRISPR spacer to be inserted into the BsaI cut plasmid CRISPR array: a ‘‘top’’ strand oligo with sequence 5’-AAAC-(30 bp spacer)-G-3’ and a ‘‘bottom’’ strand oligo with sequence 5’-AAAAC-(30 bp spacer reverse complement)-3’. For oligo cloning of the type I-E spacer *spc9R* (and others) into pACYC184-TypeIE*spcNT*, the top strand oligo had sequence 5’-ACCG-(32 bp spacer)-3’ and the bottom strand oligo had sequence 5’-ACTC-(32 bp spacer reverse complement)-3’. The two oligos were phosphorylated with T4 polynucleotide kinase (NEB, M0201S) in a 50 μL reaction: 1.5 μL 100 μM top oligo, 1.5 μL 100 μM bottom oligo, 41 μL nuclease-free water, 5 μL T4 DNA ligase reaction buffer (NEB, B0202S), 1 μL T4 polynucleotide kinase (NEB, M0201S). The reaction was incubated at 37°C for 1 hour in a thermal cycler. After phosphorylation, oligos were annealed by adding 2.5 μL of 1 M sodium chloride (Fisher Scientific, S271-3) solution to the 50 μL reaction and incubating for 5 minutes at 98°C and then allowing the reaction to gradually cool to room temperature (approximately 2 hours). The annealed oligos were diluted 1:10 in nuclease-free water and ligated into the BsaI-digested plasmid in a 20 μL reaction: 10 μL BsaI-digested plasmid, 6 μL nuclease-free water, 1 μL 1:10 diluted annealed oligos, 5 μL T4 DNA ligase reaction buffer (NEB, B0202S), 1 μL T4 DNA ligase (NEB, M0202M). The ligation reaction was performed at room temperature overnight. The next day, 5 μL of the ligation reaction was transformed

into 50 μL of rubidium chloride chemically competent *E. coli* K-12 MG1655 cells or chemically competent *E. coli* EC100 cells. The CRISPR spacers used for λ Red experiments, cloned by BsaI cloning described here, are listed in **Table 9.4**.

9.5 Strain construction

Transductions with bacteriophage P1 (ATCC 25404-B1) were performed to move the *polB*, *dinB* and *umuC* kanamycin-marked deletions in Keio collection strains into the KD263 strain expressing the λ_*L6-R* spacer. λ_*L6-R* cells or variants thereof were grown in 5 mL LB overnight at 37°C. Overnight cultures were resuspended in a 1/2X volume of P1 salts solution: 10 mM $CaCl_2$ (Fisher Scientific, BP510-500), 5 mM $MgSO_4$ (Sigma-Aldrich, M1880-500G). 100 mL cells were mixed with several 10-fold dilutions of a fresh, high-titer P1 stock (10^9-10^{10} PFU/mL) and incubated at room temperature for 30 minutes. 1 mL LB + 200 mL 1 M sodium citrate were added to each tube and cells were incubated at 37°C for 1 hour with shaking. Cells were pelleted, resuspended in 50 mL LB and spread onto LB/kanamycin plates. Colonies from the plates that received the lowest amount of P1 phage were re-struck to single colonies to ensure phage removal. Colonies were checked for the presence of the *polB*, *dinB* and *umuC* deletions by PCR with primers JW3077 and JW2096, JW3077 and JW2097 and JW3077 and JW2099 respectively, and the identity of the spacer within the CRISPR array was confirmed by amplification with primers JW3100 and JW3101 followed by Sanger sequencing.

λ Red recombineering was used to generate the *E. coli* K-12 BW25113 Δ*xthA* Δ*nfo* strain. An overnight culture of the *E. coli* K-12 BW25113 Δ*nfo* Keio strain carrying the pAM38(*red*) plasmid with chloramphenicol resistance was diluted and grown to OD_{600}~0.3 and then induced with 0.2% L-arabinose till OD_{600} ~1-1.2. Cells were made electrocompetent by washing twice with cold water and electroporated (1 mm Bio-Rad Gene Pulser cuvette at 1.8 kV) with a PCR product carrying a *xthA:tetR* gene replacement matching the *xthA:kanR* gene replacement found in the *E. coli* K-12 BW25113 Δ*xthA* Keio strain, with ~50 bp homology upstream and downstream of the *xthA* locus in the PCR product. After ~2 hours of recovery, cells were plated on LB agar plates with kanamycin at 50 μg/mL and tetracycline at 5 μg/mL to select for double mutants. Double knockouts were confirmed by PCR. After confirmation, strains were grown overnight in LB with kanamycin at 50 μg/mL and tetracycline at 5 μg/mL (but no chloramphenicol which selects for the plasmid) and with 0.2% L-arabinose induction. Without antibiotic selection, induced plasmid was rapidly lost due to toxicity from λ Red overexpression. Strains were frozen and struck out on appropriate antibiotic plates to confirm both double knockouts and loss of the recombineering plasmid.

9.6 Preparation of phage λ*vir* parental stock

λ*vir* was obtained from Bruce Levin and frozen down at -80°C: 900 μL phage in LB medium with 100 μL dimethyl sulfoxide (Sigma-Aldrich, D2650-100ML). A pipette tip was used to scrape off a tiny portion of frozen phage stock and resuspended in 20

μL LB medium. Serial dilutions of phage stock were prepared from the resuspended phage and spotted on a fresh LB top agar (Invitrogen LB broth base, Thermo Fisher Scientific, 12780-029; 0.5% final concentration of Fisher Bioreagents agar, Fisher Scientific, BP1423-500) lawn of *E. coli* K-12 MG1655 Δ9, in LB agar supplemented with 10 mM $MgSO_4$ (Sigma-Aldrich, M1880-500G). The plate was incubated at 37°C overnight after drying at room temperature for 25 minutes. The next day a single phage plaque was picked from the top agar lawn using a P20 pipette set to 15 μL and resuspended in 20 μL LB medium. 5 μL of resuspended phage was spotted on to a fresh top agar lawn of *E. coli* K-12 MG1655 Δ9 in LB agar supplemented with 10 mM $MgSO_4$. The next day the phage spot on the top agar lawn was picked with a P20 pipette set to 15 μL and resuspended in 100 μL of an overnight stationary culture of *E. coli* K-12 MG1655 Δ9. The resuspended phage culture was mixed with 6 mL top agar supplemented with 10 mM $MgSO_4$ and poured over a LB agar plate. This was repeated twice for a total of 3 top agar plates. The plates were incubated at 37°C overnight after drying at room temperature for 25 minutes. The next day top agar from each plate was scraped off with 5 mL LB medium into a single 50 mL conical tube. The tube was then spun down at 15,000 x *g* for 10 minutes at 4°C. The phage-containing supernatant was filtered using an Acrodisc 13 mm SUPOR 0.45 μm syringe filter (Pall, 4604) into a 15 mL conical tube. The phage stock was stored at 4°C.

9.7 Construction of λ*vir* wild-type and mutant phage stocks

The λ*vir* parental stock was used to construct wild-type and mutant phage stocks. 1 μL of the phage stock was resuspended in 100 μL of an overnight stationary culture of *E. coli* K-12 MG1655 Δ9 containing a recombinant pUT18C-based plasmid. Each recombinant pUT18C-based plasmid contained a cloned segment of phage λ*vir* DNA with the desired gene or genomic region modified or deleted. The resuspended phage culture was mixed with 6 mL LB top agar (Invitrogen LB broth base, Thermo Fisher Scientific, 12780-029; 0.5% final concentration of Fisher Bioreagents agar, Fisher Scientific, BP1423-500) supplemented with 100 μg/mL carbenicillin (GoldBio, C-103-25) and 10 mM $MgSO_4$ (Sigma-Aldrich, M1880-500G) and poured over a LB agar plate supplemented with 100 μg/mL carbenicillin. Plates were incubated at 37°C overnight after drying at room temperature for 25 minutes. The next day top agar from each plate was scraped off with 5 mL LB medium into 50 mL conical tubes. Tubes were spun down at 15,000 x *g* for 10 minutes at 4°C. Recombinant phage-containing supernatant was filtered using an Acrodisc 13 mm SUPOR 0.45 μm syringe filter (Pall, 4604) into a 15 mL conical tube.

Serial dilutions of recombinant phage were prepared and spotted on a fresh top agar lawn of *E. coli* K-12 MG1655 Δ9 containing a pCas9 plasmid in LB agar supplemented with 25 μg/mL chloramphenicol (GoldBio, C-105-100) and 10 mM $MgSO_4$ (Sigma-Aldrich, M1880-500G). Each pCas9 plasmid carried a type II-A

CRISPR spacer targeting the phage region that was modified or deleted to select specifically for recombinant phage with the desired deletion or modification. Top agar plates were incubated at 37°C overnight after drying at room temperature for 25 minutes. For each recombinant phage, the next day a single phage plaque was picked from the top agar lawn using a P20 pipette set to 15 μL and resuspended in 20 μL LB medium. Serial dilutions of the resuspended phage were prepared and spotted on a fresh top agar lawn of *E. coli* K-12 MG1655 Δ9 containing the corresponding pCas9 plasmid in LB agar supplemented with 25 μg/mL chloramphenicol and 10 mM $MgSO_4$. The top agar plate was incubated at 37°C overnight after drying at room temperature for 25 minutes. The next day a single phage plaque was picked from the top agar lawn using a P20 pipette set to 15 μL and resuspended in 20 μL LB medium. 5 μL of resuspended phage was spotted on to a fresh top agar lawn of *E. coli* K-12 MG1655 Δ9 containing the corresponding pCas9 plasmid in LB agar supplemented with 25 μg/mL chloramphenicol and 10 mM $MgSO_4$. The top agar plate was incubated at 37°C overnight after drying at room temperature for 25 minutes. The next day the phage spot on the top agar lawn was picked with a P20 pipette set to 15 μL and resuspended in 100 μL of an overnight stationary culture of *E. coli* K-12 MG1655 Δ9 containing the corresponding pCas9 plasmid. The resuspended phage culture was mixed with 6 mL top agar supplemented with 25 μg/mL chloramphenicol and 10 mM $MgSO_4$ and poured over a LB agar plate. This was repeated twice for a total of 3 top agar plates. The plates were incubated at 37°C overnight after drying at room temperature for 25 minutes.

The next day top agar from each plate was scraped off with 5 mL LB medium into a single 50 mL conical tube. The tube was then spun down at 15,000 x *g* for 10 minutes at 4°C. The phage-containing supernatant was filtered using an Acrodisc 13 mm SUPOR 0.45 μm syringe filter (Pall, 4604) into a 15 mL conical tube. All phage stocks were stored at 4°C.

λ*vir* Δ*gam* was generated by passaging λ*vir* through cells harboring pUT18C-D*gam* and plaquing the lysate on a lawn of cells harboring pCas9:JW1370-1. λ*vir* Δ*exo* was generated by passaging λ*vir* through cells harboring pUT18C-D*exo* and plaquing the lysate on a lawn of cells harboring pCas9:*spc12*. λ*vir* Δ*bet* was generated by passaging λ*vir* through cells harboring pUT18C-D*bet* and plaquing the lysate on a lawn of cells harboring pCas9:JW1552-3. λ*vir* Δ*exo*Δ*bet* was generated by passaging λ*vir* through cells harboring pUT18C-D*exo*D*bet* and plaquing the lysate on a lawn of cells harboring pCas9:JW1552-3. λ*vir* Δ*red* was generated by passaging λ*vir* through cells harboring pUT18C-D*red* and plaquing the lysate on a lawn of cells harboring pCas9:JW1552-3. In each case, *red* gene mutations were confirmed within an escaper plaque by PCR and Sanger sequencing. Wild-type λ*vir* was generated by passaging λ*vir* through cells harboring pUT18C and plaquing the lysate on a lawn of cells harboring pCas9:*spcNT*.

For phages containing *chi* site modifications (λ*vir chi*^{1} and λ*vir chi*$^{2\text{-}7}$), *chi*-containing phages were first generated using the method described above. The

process was then repeated for *gam*, *exo*, *bet*, *exo-bet* and *red* gene deletions for *chi*-containing phages. For wild-type λ*vir* phage stocks, all steps were mirrored to mutant phage generation, except instead a pUT18C empty plasmid was used in recombineering and pCas9::*spcNT* was used in subsequent selection steps. λ*vir chi*^{1} was generated by passaging λ*vir* through cells harboring pUT18C-*chi*D and plaquing the lysate on a lawn of cells harboring pCas9:JW1546-7. The *chi*^{1} (*chi*D) mutation was confirmed within an escaper plaque by PCR and Sanger sequencing. λ*vir chi*$^{2-7}$ was generated by passaging λ*vir* through cells harboring pUT18C-3*chi*F+R and plaquing the lysate on a lawn containing both cells harboring pCas9:JW1558-9 and cells harboring pCas9:JW1401-2. The *chi*$^{2-4}$ (*chi*3F) and *chi*$^{5-7}$ (*chi*3R) mutations were individually confirmed within an escaper plaque by PCR and Sanger sequencing.

Final phage stocks were PCR checked and Sanger sequenced to confirm appropriate gene deletions or genetic modifications. Furthermore, serial dilutions of each phage stock were prepared and 3.5 μL dilutions were spotted and dripped on fresh top agar lawns of *E. coli* K-12 MG1655 on LB agar supplemented with 10 mM $MgSO_4$. A series of four separate dilutions were prepared for each phage stock on top agar lawns to accurately determine phage titers. Top agar plates were incubated at 37°C overnight after drying at room temperature for 25 minutes. The next day, phage plaques were counted to determine titers. Additionally, for *gam*, *exo*, *bet*, *exo-bet* and *red* gene deletion mutants, 16 individual plaques were isolated per phage

stock and PCR checked to confirm that they all contained the appropriate gene deletion(s). The phages used in this thesis are listed in **Table 9.5**.

The replication-deficient phage λΔ*P* was generated as a lysogen using λ Red recombineering, facilitated by the plasmid pKOBEG-A (AmpR) which expresses *gam*, *bet*, and *exo* from the arabinose-inducible pBAD promoter (Chaveroche et al., 2000). MG1655 cells harboring a λ lysogen (obtained from J.W. Roberts) and the pKOBEG-A plasmid were grown at 30°C overnight in LB supplemented with 100 μg/mL carbenicillin. The overnight culture was diluted 1:50 into 500 mL fresh media and grown until OD_{600} reached 0.2. The culture was supplemented with 0.2% L-arabinose (Sigma-Aldrich, A91906-100G-A) and grown further to an OD_{600} of 1. Cells were then centrifuged at 4000 x *g* for 10 mins at 4°C and washed twice in 500 mL sterile ice cold water. After the second wash, cells were centrifuged again and resuspended in 40 mL sterile ice cold water. After a final centrifugation, cells were resuspended in 1.5 mL cold 10% glycerol (Fisher Scientific, G33-500) and 60 μL aliquots were stored at -80°C.

To generate a kanamycin-marked deletion construct for the λ replication gene P (λΔP), the kanamycin resistance gene on pKD4 (Datsenko and Wanner, 2000) was amplified using primers JW1518 and JW1519. This amplicon was then re-amplified with primers JW1520 and JW1521, introducing 50 bp of homology to the λ regions upstream and downstream of P. The second amplicon (~100 ng) was electroporated

into electrocompetent, arabinose-induced MG1655::λ cells harboring pKOBEG-A and transformants were selected on LB-kanamycin plates grown at 30°C. Transformants were re-struck on LB-kanamycin plates grown at 42°C to promote pKOBEG-A plasmid loss. Finally, clones were patched onto a LB-carbenicillin plate to ensure loss of pKOBEG-A.

To generate the λΔP phage stock, MG1655 cells harboring the λΔP lysogen and the plasmid pCL1920-P were grown into exponential phase and treated with 0.1 μg/mL mitomycin C (AG Scientific, M-1108) for 2 hours. Phage-containing supernatants were filtered using Acrodisc 13 mm SUPOR 0.45 μm syringe filters (Pall, 4604) and used to isolate single plaques on a lawn of MG1655 cells harboring pCL1920-P. Single plaques were used to lyse plates of MG1655 cells harboring pCL1920-P and plate lysates were pooled to create λΔP stocks.

9.8 Phage λ*vir* genome sequencing and assembly

λ*vir* phage capsids were digested with 50 μg/mL proteinase K and phage genomic DNA was extracted using the DNeasy Blood & Tissue kit (QIAGEN, 69504). Genomic DNA was sequenced following a previously described method in our laboratory (Meeske et al., 2019). Briefly, isolated DNA was sheared using a pre-split snap-cap 6x16 mm Covaris microTUBE (Covaris, 520045) in a Covaris S220 focused-ultrasonicator and prepared for next generation sequencing using the Illumina TruSeq Nano DNA LT kit (Illumina, 20015964). Paired-end 2 × 75-bp

sequencing was conducted using the 150-cycle MiSeq Reagent Kit v3 (Illumina, MS-102-3001) on the Illumina MiSeq platform. Reads were quality-trimmed using Sickle (https://github.com/najoshi/sickle) and assembled into contigs using ABySS (Simpson et al., 2009) (https://github.com/bcgsc/abyss). Finally, contigs were mapped to a reference phage λ genome (GenBank: KT232076.1) using Medusa (Bosi et al., 2015) (http://combo.dbe.unifi.it/medusa). Automated genome annotation was performed using SnapGene and the reference phage λ genome (GenBank: KT232076.1).

9.9 qPCR of phage λ*vir* DNA replication

To quantify DNA replication of phage λ*vir* within an infected *E. coli* cell, an overnight culture of *E. coli* K-12 MG1655 cells carrying pCas9::*spcNT* (non-targeting spacer) was diluted 1:50 in 50 mL of LB medium (BD Difco LB Broth, Lennox, BD 240220) supplemented with 25 μg/mL chloramphenicol (GoldBio, C-105-100) and 10 mM $MgSO_4$ (Sigma-Aldrich, M1880-500G). After 1 hour of growth, OD_{600} was measured and the culture was normalized to OD_{600} = 0.3. 700 μL of culture was dispensed between multiple 1.5 mL Eppendorf tubes, corresponding to three replicates and two timepoints (15 and 30 minutes) for each phage being monitored. These 700 μL cultures were infected with corresponding λ*vir* phages at MOI 1 and incubated at 37°C with shaking for either 15 or 30 minutes. At each timepoint, samples were removed from the incubator and tubes spun down at 15,000 rpm for 2 minutes in a tabletop microcentrifuge at 4°C. Supernatants were removed and cell pellets immediately frozen down at -80°C for DNA extraction later. Additionally, three

uninfected tubes were also prepared for DNA extraction as no-phage controls for qPCR.

Total DNA was extracted from frozen *E. coli* cell pellets using the Promega Wizard Genomic DNA Purification Kit (Promega, A1125) following the protocol for Gram-negative bacteria. Extracted DNA was quantified using the Qubit 1X dsDNA HS Assay Kit (Thermo Fisher Scientific, Q33231) and each sample was normalized to 4 ng/μL or 5 ng/μL. A total of 25 ng DNA was used as input for qPCR, performed using Applied Biosystems Fast SYBR Green Master Mix (Thermo Fisher Scientific, 4385612) on the Applied Biosystems QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific, A28567) with primer pairs AA385/AA386 (λ*vir* DNA) and AA387/AA388 (*E. coli* K-12 MG1655 *dxs* control). The sequences of the primers are listed in **Table 9.1**.

9.10 Plaque assays of phage λ*vir* in *E. coli* K-12 MG1655

Overnight cultures were launched from single colonies in 3 mL of LB medium (BD Difco LB Broth, Lennox, BD 240220) with appropriate antibiotic(s). Top agar lawns of *E. coli* were prepared by mixing 100 μL of overnight culture with 6 mL of LB top agar (Invitrogen LB broth base, Thermo Fisher Scientific, 12780-029; 0.5% final concentration of Fisher Bioreagents agar, Fisher Scientific, BP1423-500) supplemented with appropriate antibiotic(s) and 10 mM $MgSO_4$ (Sigma-Aldrich, M1880-500G). Top agar mixtures were poured over LB agar (BD Difco LB Agar, Miller, BD 244510) in 10 cm plates supplemented with appropriate antibiotic(s).

Where necessary, 0.2% L-arabinose (Sigma-Aldrich, A91906-100G-A) or 1 mM IPTG (GoldBio, I2481C100) was included in the LB top agar and the LB agar plate. Plates were dried at room temperature, partially open by a sterilizing flame, for 25 minutes for the top agar to solidify. Serial dilutions of phage stock were prepared and spotted on the top agar after drying. Before making serial dilutions, aliquots of phage stocks were uniformly normalized to 1×10^7 plaque forming units/uL (PFU/μL) for all efficiency of plaquing and plaque formation assays, except **Fig. 2.4** where phages were uniformly normalized to 5×10^6 PFU/μL before making serial dilutions for plaquing. For imaging of plaque assays, 2.5 μL of each phage dilution was spotted on top agar using a multichannel pipette. For quantification of phage titers (and in some cases, imaging of plaque assays) and isolation of single phage plaques for phage DNA sequencing, 3.5 μL of each phage dilution was spotted on top agar using a multichannel pipette and the plate was tilted to allow phage spots to drip down the plate for easier quantification and isolation of single plaques. In either case, plates were incubated at 37°C overnight (30°C in certain cases) after drying at room temperature for 25 minutes or until the plates were completely dry.

9.11 Plaque assays in *E. coli* KD263 and KD263-derived strains

Overnight cultures were launched from single colonies in 3 mL of LB medium (BD Difco LB Broth, Lennox, BD 240220). For experiments with KD263-derived deletion mutants, 50 μg/mL kanamycin (GoldBio, K-120-100) was included in overnight LB liquid media (BD Difco LB Broth, Lennox, BD 240220), LB top agar

(Invitrogen LB broth base, Thermo Fisher Scientific, 12780-029; 0.5% final concentration of Fisher Bioreagents agar, Fisher Scientific, BP1423-500) and LB agar (BD Difco LB Agar, Miller, BD 244510) plates to select for the mutant strains. For imaging of plaque assays, top agar lawns of *E. coli* were prepared by mixing 100 μL of overnight culture with 6 mL of LB top agar supplemented with 10 mM $MgSO_4$ (Sigma-Aldrich, M1880-500G). Top agar mixtures were poured over LB agar in 10 cm plates. Where necessary, 1 mM L-arabinose (Sigma-Aldrich, A91906-100G-A) for induction of Cascade, Cas1, Cas2 and type I-E spacers and 1mM IPTG (GoldBio, I2481C100) for induction of Cas3 were included in the LB top agar and the LB agar plate. Plates were dried at room temperature, partially open by a sterilizing flame, for 25 minutes for the top agar to solidify. Serial dilutions of phage stock, normalized to 1×10^7 PFU/μL, were prepared and spotted on the top agar after drying. Before making serial dilutions, aliquots of phage stocks were uniformly normalized to 1×10^7 PFU/uL. 3.5 μL of each phage dilution was spotted on top agar using a multichannel pipette and the plate was tilted to allow phage spots to drip down the plate for clearer visualization of small individual phage plaques. Plates were incubated at 37°C overnight after drying at room temperature for 25 minutes or until the plates were completely dry. For quantification of phage titers (and isolation of single phage plaques for phage DNA sequencing) on bacterial lawns of λ_*L1-R*, λ_*L4-R* and λ_*L6-R* strains, top agar lawns of *E. coli* were prepared by mixing 100 μL of overnight culture with a prespecified amount of phage PFUs in 6 mL of LB top agar (so as to obtain countable single plaques) supplemented with 10 mM $MgSO_4$, 1 mM L-

arabinose and 1 mM IPTG (1 mM L-arabinose and 1mM IPTG were also included in the LB agar plate). 50 μg/mL kanamycin was included in all the media for λ_*L6-R* polymerase mutant strains. The following phage PFUs were added in the top agar lawns: 5×10^5 λ*vir chi*^{1} PFUs or 5×10^7 λ*vir chi*^{1} Δ*red* PFUs for λ_*L1-R*, 5×10^6 λ*vir chi*^{1} PFUs or 2×10^8 λ*vir chi*^{1} Δ*red* PFUs for λ_*L4-R*, 1×10^5 λ*vir chi*^{1} PFUs or 1×10^7 λ*vir chi*^{1} Δ*red* PFUs for λ_*L6-R* strains. Plates were incubated at 37°C overnight after drying at room temperature for 25 minutes. For λ_*L6-R* strains in the polymerase mutant plaquing assays, CRISPR non-targeting PFUs were also determined on top agar lawns of wild-type or polymerase mutant λ_*L6-R* cells by dripping serial dilutions of λ*vir chi*^{1} phage stock on top agar lawns without any inducer. For non-targeting KD263 and λ_*E4-R* strains, plaques were dripped on lawns (with 1 mM L-arabinose and 1mM IPTG) as described above and used to determine phage titers. For infection of the λ_*L4-R* strain with λ*vir chi*^{1} Δ*red*, only 1-2 pickable plaques were obtained per infected lawn. To obtain enough escaper phages for characterization of escape mutations, 8 plates of top agar infections were prepared for picking 8 escape plaques for subsequent escaper sequencing.

9.12 Imaging and quantification of phage λ*vir* plaque assays

Overnight plaque formation assays were imaged the next day (~16-24 hours after infection) using the FluorChem HD2 system (ProteinSimple). Plaque assay images were all equally adjusted for brightness and contrast using Adobe Photoshop. Images of plaque formation assays with 3.5 μL drips were manually counted for

plaque forming units (PFUs) using Fiji (ImageJ) (https://imagej.net/software/fiji/). Three biological replicates, corresponding to overnight cultures launched from three separate single bacterial colonies, were counted for each strain and phage combination. For quantification of plaques on λ_*L1-R*, λ_*L4-R* and λ_*L6-R* strains, individual plaques were obtained across the entire bacterial lawn (see above) and counted across the entire lawn to determine PFUs. Lawn images were all equally adjusted for brightness and contrast using Adobe Photoshop and single plaques across the entire plate were manually counted using Fiji. These plaque counts were then divided by the total number of plaques used to infect each lawn to calculate the efficiency of plaquing.

9.13 Efficiency of plaquing analysis

For the plaque assays described above, efficiency of plaquing is calculated as the plaques formed by the phage on a CRISPR targeting lawn divided by the plaques formed by the same phage on a CRISPR non-targeting lawn (*spcNT* or no inducer with type I-E *spc9R*). For quantification of plaques on λ_*L1-R*, λ_*L4-R* and λ_*L6-R* strains, individual plaques were obtained across the entire bacterial lawn with 1 mM L-arabinose (Sigma-Aldrich, A91906-100G-A) and 1 mM IPTG (GoldBio, I2481C100) and counted across the entire lawn to determine plaque forming units. These plaque counts were then divided by the total number of plaques used to infect each lawn to calculate the efficiency of plaquing. For quantification of plaques for the λ_*L6-R* polymerase mutants' plaque assays, individual plaques were obtained across the

entire bacterial lawn with 1 mM L-arabinose and 1 mM IPTG and counted across the entire lawn to determine plaque forming units. These plaque counts were then divided by the total number of plaques used to infect each lawn to determine the efficiency of plaquing. Plaques were also counted from drips of serial phage dilutions on lawns of wild-type or polymerase null mutants with no inducer (CRISPR non-targeting) and the total plaques formed on these non-targeting lawns of wild-type or mutant cells were used to calculate an adjusted efficiency of plaquing for each *spcL6-R* strain. For the efficiency of plaquing bar graphs shown in Chapters 2 and 3, error bars represent the standard error of the mean of three biological replicates. Statistical analyses of plaque assays were performed using GraphPad Prism 9. Where necessary, a *t*-test was used to compare means. In each case, an unpaired parametric *t*-test was performed with the assumption that samples come from populations with a Gaussian distribution and the same standard deviation.

9.14 Target sequencing of phage λ*vir* CRISPR escapers

Six to sixteen phage plaques were isolated and resuspended in 20 μL of LB medium. 2-5 μL of each resuspended plaque was spotted on to a fresh top agar lawn of the original CRISPR targeting strain (from which the escapers were isolated) on LB agar supplemented with appropriate antibiotic and 10 mM $MgSO_4$ (Sigma-Aldrich, A91906-100G-A). For type I-E *spc9R* escapers, 0.2% L-arabinose (Sigma-Aldrich, A91906-100G-A) was included in the top agar and the LB agar plate for induction of Cascade, Cas3 and the plasmid-encoded type I-E spacer. For escapers of KD263-

derived λ_*L1-R*, λ_*L4-R* and λ_*L6-R* strains, 1 mM L-arabinose (Sigma-Aldrich, A91906-100G-A) and 1 mM IPTG (GoldBio, I2481C100) were included in the top agar and the LB agar plate for induction of Cascade, Cas1, Cas2, Cas3 and type I-E spacers. Top agar plates were incubated at 37°C overnight after drying at room temperature for 25 minutes or until completely dry. The next day phage escaper spots were picked using a 20 μL pipette set to 15 μL and resuspended in 20 μL of colony lysis buffer (Pyenson et al., 2017). Resuspended phage mixtures were boiled at 98°C for 10 minutes in a thermal cycler, and 1-3 μL of the boiled phage mixture was then used as template for PCR amplification with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, F530L). The oligonucleotide primers used for protospacer target sequencing are listed in **Table 9.1**. PCR products were submitted to Sanger sequencing by Genewiz. For escapers of *spc14*, isolated escaper plaques were resuspended in 20 μL of LB. Serial dilutions of each resuspended escaper plaque were prepared and then spotted and dripped on fresh top agar lawns of the original *spc14* targeting *E. coli* in LB agar supplemented with 25 μg/mL chloramphenicol and 10 mM $MgSO_4$. The next day single escaper plaques were picked for each original escaper phage and resuspended in 20 μL of colony lysis buffer for use in PCR.

9.15 In vivo CRISPR-Cas9 cleavage of viral DNA

To observe CRISPR-Cas9 cleavage of anti-viral targets, overnight cultures of *E. coli* K-12 MG1655 cells carrying pCas9::*spcNT* (non-targeting) or pCas9::*spc9*

(targeting protospacer 9) were diluted to an OD_{600} of ~0.05 in LB medium supplemented with 25 μg/mL chloramphenicol (GoldBio, C-105-100) and 10 mM $MgSO_4$ (Sigma-Aldrich, A91906-100G-A). After 1 hour and 10 minutes of growth, OD_{600} was measured for each culture, and each sample was normalized to OD_{600} = 0.25. Cultures were then infected at MOI 5 with λ*vir chi*^{1} or λ*vir chi*^{1} Δ*gam* for 25 minutes prior to centrifugation and flash-freezing of cell pellets. All samples were stored at -80°C until ready for genomic DNA purification using the DNeasy Blood & Tissue Kit (QIAGEN, 69504) following the protocol for Gram-negative organisms. Purified genomic DNA was sheared using a pre-split snap-cap 6x16 mm Covaris microTUBE (Covaris, 520045) in a Covaris S220 focused-ultrasonicator and prepared for next generation sequencing using the Illumina TruSeq Nano DNA LT kit (Illumina, 20015964). Paired-end 2 × 75-bp sequencing was conducted using the 150-cycle MiSeq Reagent Kit v3 (Illumina, MS-102-3001) on the Illumina MiSeq platform. Illumina paired-end sequencing reads were aligned to phage genomes using the Burrows-Wheeler Aligner (Li and Durbin, 2009). The resulting .sam files were parsed and visualized with custom Python scripts. For *in vivo* cleavage assays of pCas9::*spcNT*, pCas9::*spc45* or pCas9::*spc45c* with λ*vir chi*^{1}, the procedure was carried out as described above except that genomic DNA was isolated from frozen *E. coli* cell pellets using the Promega Wizard Genomic DNA Purification Kit (Promega, A1125) following the protocol for Gram-negative bacteria.

9.16 Competition experiment between λ*vir chi*^{1} and λvir *chi*^{1} Δ*red*

Overnight cultures of *E. coli* K-12 MG1655 cells carrying pCas9::*spcNT* (non-targeting spacer), pCas9::*spc9* or pCas9::*spc45* were launched from single colonies in 3 mL of LB medium with 25 μg/mL chloramphenicol (GoldBio, C-105-100). The next day, each overnight culture was normalized to OD_{600} = 3.5. Top agar lawns of *E. coli* were prepared by mixing 100 μL of the normalized overnight culture with a 1:1 phage PFU mixture of λ*vir chi*^{1} and λ*vir chi*^{1} Δ*red* at a total MOI of 20 in 6 mL of LB top agar supplemented with 25 μg/mL chloramphenicol and 10 mM $MgSO_4$ (Sigma-Aldrich, A91906-100G-A). The plates were incubated at 37°C overnight after drying at room temperature for 25 minutes. The phage infection mix was stored at 4°C overnight. The next day top agar from each infection plate was scraped off with 5 mL LB medium into a single 50 mL conical tube. The tube was then spun down at 15,000 x *g* for 10 minutes at 4°C. The phage-containing supernatant was filtered using an Acrodisc 13 mm SUPOR 0.45 μm syringe filter (Pall, 4604) into a 15 mL conical tube. The stored phage infection mix and the filtered phage supernatants were then each used to prepare serial phage dilutions to obtain single phage plaques (for PCR of the λ *red* locus) on fresh top agar lawns of *E. coli* K-12 MG1655 cells carrying pCas9::*spcNT* in LB agar supplemented with 25 μg/mL chloramphenicol and 10 mM $MgSO_4$. Top agar plates were incubated at 37°C overnight after drying at room temperature for 25 minutes. The next day, individual phage plaques were picked and resuspended in 20 μL of colony lysis buffer (Pyenson et al., 2017). Resuspended phage mixtures were boiled at 98°C for 10 minutes in a thermal cycler, and 0.5 μL of

the boiled phage mixture was then used as template for PCR amplification of the λ *red* locus. PCR was performed in a 20 μL total reaction volume using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, F530L) and 1 μM each primer (AA549/AA550) with an annealing temperature of 71°C and an extension time of 1.5 minutes. For each set of phage plaques, a total of 12 phage plaques were first isolated and PCR amplified with AA549/AA550 (**Table 9.1**). PCR products were run on a 1.5% agarose gel at 140V for 35 minutes to determine the presence (~2.3 kb) or absence (~0.6 kb) of the phage *red* operon. These were then counted to determine the total number of *red*+ (λ*vir chi*^{1}) and *red*- (λ*vir chi*^{1} Δ*red*) phages. Only plaques that showed a single PCR band on the agarose gel were used in our analyses. In cases where the agarose gel showed both *red*+ and *red*- PCR bands for an isolated phage plaque, a new plaque was selected randomly from the serial dilution top agar lawns, PCR amplified with AA549/AA550 and run on an agarose gel as before. The competition experiment was performed a total of three times for three biological replicates. Additionally, the twelve phage plaques that were isolated for each run of the *spc45*-targeting competition for PCR of the *red* locus were also checked for potential target mutations in the protospacer through PCR of the *spc45* target region using primers AA138/AA139 (**Table 9.1**) and subsequent Sanger sequencing by Genewiz with primer AA164 (**Table 9.1**).

9.17 Next-generation sequencing of *spc9* escape phages

Overnight cultures of *E. coli* K-12 MG1655 cells or *E. coli* K-12 MG1655 Δ*dinB* cells carrying pCas9::*spc9* were launched from single colonies in 3 mL LB medium with appropriate antibiotic(s). Serial dilutions of phage stock (λ*vir chi*^{1} or λ*vir chi*^{1} Δ*red*) were prepared and spotted on fresh top agar lawns of *E. coli* K-12 MG1655 carrying pCas9::*spc9* supplemented with 25 μg/mL chloramphenicol (GoldBio, C-105-100) and 10 mM $MgSO_4$ (Sigma-Aldrich, A91906-100G-A). In each case, 3.5 μL of each phage dilution was spotted on top agar using a multichannel pipette and the plate was tilted to allow phage spots to drip down the plate. Plates were incubated at 37°C overnight after drying at room temperature for 25 minutes or until the plates were completely dry. With pKM208 plasmid-containing cells, 100 μg/mL carbenicillin (GoldBio, C-103-25) and 1mM IPTG (GoldBio, I2481C100) were also included in the LB top agar and the LB agar plate, and cultures and plates were incubated at 30°C.

The next day, top agar phage drips containing ~100-1000 plaques (one full drip dilution per top agar plate) were scooped up using an inverted sterile pipette tip and resuspended in 500 μL LB medium in 1.5 mL Eppendorf tubes. Tubes were spun down at 15,000 rpm for 3 minutes in a tabletop microcentrifuge at 4°C. Phage-containing supernatants were then filtered using Acrodisc 13 mm SUPOR 0.45 μm syringe filters (Pall, 4604) into fresh 1.5 mL Eppendorf tubes. To isolate phage DNA as template for PCR, 5 μL of phage supernatant was mixed with 15 μL colony lysis buffer (Pyenson et al., 2017) and boiled at 98°C for 10 minutes in a thermal cycler.

PCR of the *spc9* target was performed with 5 μL phage lysis input in a 50 μL PCR reaction using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, F530L) and barcoded primers with identical 4-9 bp DNA barcodes on forward and reverse primers. The barcoded oligonucleotide primers used for PCR are listed in Table S3. PCR products were run on a 2% agarose gel at 140 V for 25 minutes. Gel-run PCR products were then gel extracted using the QIAquick Gel Extraction Kit (QIAGEN, 28704). Extracted DNA was quantified using the Qubit 1X dsDNA HS Assay Kit (Thermo Fisher Scientific, Q33231) and analyzed for DNA fragment size using the Agilent 2200 TapeStation system with the High Sensitivity D1000 ScreenTape (Agilent, 5067-5584 and 5067-5585). Extracted samples were then pooled together for a total DNA concentration of 10-20 ng/μL, with each sample at equal ng/μL concentrations. The pooled PCR sample was prepared for next generation sequencing using the Illumina TruSeq Nano DNA LT kit (Illumina, 20015964), followed by paired-end 2 × 75-bp sequencing using the 150-cycle MiSeq Reagent Kit v3 (Illumina, MS-102-3001) on the Illumina MiSeq platform.

9.18 Next-generation sequencing data analysis of *spc9* escape phages

Forward and reverse sequencing reads of *spc9* target PCR products were extracted from raw MiSeq FASTQ files and quality filtered with a Phred quality cutoff of 10 using Python v3.8 on PyCharm CE. For reads passing quality filter, corresponding forward and reverse reads were compiled into a text file for

downstream data analysis using a custom Python script. Reads were first organized into individual lists according to their primer barcodes. Reads were selected only if both forward and reverse reads contained the same barcode to avoid issues with barcode switching between samples during library preparation for next generation sequencing. Only sequencing reads from the bottom strand 5' end of the PCR products were selected since only these contained the full *spc9* protospacer and PAM sequence. Barcode and reverse PCR primer sequences were subtracted from each bottom strand read. Each read was then mapped to the *spc9* protospacer region reverse complement reference sequence and base pair mismatches between the reference sequence and each sequencing read was recorded and enumerated. For each sample barcode, a list of mutations was generated with the following information: bp position in protospacer/PAM, wild-type base, mutated base, fraction of total mutations (count of the specific mutation divided by the total number of mutations recorded), normalized mutation count (count of the specific mutation divided by the total number of sequencing reads analyzed). This information was then outputted using Python to a separate Microsoft Excel file for each sample barcode. The major escape mutations, ones with the largest normalized mutation counts (graphed as normalized mutated reads), were collected into a Prism file for further graphical and statistical analyses using GraphPad Prism 9. For the *spc9* escape mutation bar graphs in Chapter 3, the number of reads with a specific mutation is normalized to the total reads analyzed for each sample to give the values for normalized mutated reads. Statistical analyses were performed using GraphPad

Prism 9: *t*-tests were used to compare mean values for each PAM mutation. In each case, an unpaired parametric *t*-test was performed with the assumption that samples come from populations with a Gaussian distribution and the same standard deviation.

9.19 Liquid culture time course of phage propagation

An overnight culture of *E. coli* K-12 MG1655 cells carrying pCas9::*spcNT* (non-targeting spacer) was diluted 1:50 in 10 mL LB medium supplemented with 25 μg/mL chloramphenicol (GoldBio, C-105-100) and 10 mM $MgSO_4$ (Sigma-Aldrich, A91906-100G-A). Two separate starter cultures were launched, for infection with two phages: λ*vir chi*^{1} and λ*vir chi*^{1} Δ*red*. After one hour of growth, OD_{600} was measured for each culture, and each sample was normalized to OD_{600} = 0.3. Cultures were then infected at MOI 0.1 with each phage and incubated for 5 minutes at 37°C with shaking. After 5 minutes, cultures were removed from the incubator and spun down at 4300 x *g* for 5 minutes at 4°C. Supernatant was discarded and the pellet resuspended in 1 mL LB medium and transferred to fresh 1.5 mL Eppendorf tubes. Tubes were spun down at 8000 rpm for 1 minute in a tabletop microcentrifuge at 4°C. Supernatant was discarded and the pellet resuspended in 1 mL LB medium. Cells were washed in this manner another two times to remove unadsorbed phage. Cells were kept on ice between washes. After the final wash, pellets were resuspended in 10 mL LB medium supplemented with 25 μg/mL chloramphenicol and 10 mM $MgSO_4$. 100 μL and 500 μL of each resuspended culture was used to determine infective centers and unadsorbed phage titers respectively. Resuspended cultures were placed back in the

37°C shaking incubator for a total of 90 minutes post wash. Phage-infected cultures were sampled at the following time points: 30, 45, 60, 80, 95 minutes post wash (i.e. 35, 50, 65, 85, 100 minutes post infection). At each time point, 400 μL was collected from each phage-infected culture and spun down in 1.5 mL Eppendorf tubes at 15,000 rpm for 3 minutes in a tabletop microcentrifuge at 4°C. 100 μL of phage supernatant was collected, and phage-containing supernatants filtered using Acrodisc 13 mm SUPOR 0.45 μm syringe filters (Pall, 4604) into fresh 1.5 mL Eppendorf tubes. Filtered phage supernatants were used to prepare two sets of serial phage dilutions to estimate phage titers on fresh top agar lawns of *E. coli* K-12 MG1655 in LB agar supplemented with 10 mM $MgSO_4$. Top agar plates were incubated at 37°C overnight after drying at room temperature for 25 minutes. The next day, phage plaques were counted to determine titers.

9.20 Liquid culture time course of phage escape during *spc45* targeting

An overnight culture of *E. coli* K-12 MG1655 cells carrying pCas9::*spc45* and one carrying pCas9::*spc45c* were each diluted 1:50 in 10 mL LB medium supplemented with 25 μg/mL chloramphenicol (GoldBio, C-105-100) and 10 mM $MgSO_4$ (Sigma-Aldrich, A91906-100G-A). Two separate 10 mL starter cultures were prepared for each spacer for infection with each of two phages: λ*vir chi*^{1} and λ*vir chi*^{1} Δ*red.* After 1 hour of growth, OD_{600} was measured for each of the four outgrowth cultures, and each was normalized to OD_{600} = 0.3. The normalized 10 mL liquid

cultures of *spc45*- and *spc45c*- targeting cultures were then infected with either λ*vir chi*[1] or λ*vir chi*[1] Δ*red* at MOI 1. Infection cultures were incubated at 37°C with shaking for 9 hours. For the first 5 hours, 400 μL of culture was removed from each sample every 50 minutes and then every 2 hours for the next 4 hours for a total time course of 9 hours post infection. At each collection point, 400 μL of culture was transferred to 1.5 mL Eppendorf tubes and spun down at 15,000 rpm for 3 minutes in a tabletop microcentrifuge at 4°C. Phage-containing supernatants were then filtered using Acrodisc 13 mm SUPOR 0.45 μm syringe filters (Pall, 4604) into fresh 1.5 mL Eppendorf tubes. After completion of the time course, phage supernatants from each timepoint were used to prepare serial phage dilutions to estimate phage titers on fresh top agar lawns of *E. coli* K-12 MG1655 carrying pCas9::*spcNT* in LB top agar supplemented with 25 μg/mL chloramphenicol and 10 mM $MgSO_4$. Top agar plates were incubated at 37°C overnight after drying at room temperature for 25 minutes. The next day, phage plaques were counted to determine titers at each timepoint. Additionally, eight phage plaques were isolated for each timepoint and each infection culture to identify potential target mutations in the protospacer through PCR of the *spc45* target region using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, F530L) with primers AA138/AA139 (**Table 9.1**) and subsequent Sanger sequencing by Genewiz with primer AA164 (**Table 9.1**).

9.21 Preparation of phage stocks for Brig1 experiments

λ*vir*, T4 and T7 were generous gifts from Bruce Levin. T2, T3, T5 and T6 phages were purchased from ATCC. Phages were first grown up in 10 mL cultures of exponentially growing *E. coli* K-12 MG1655 or EC100 cells at OD_{600}~0.3. The phage-added cultures were incubated at 37°C with shaking overnight. Tubes were then spun down at 15,000 x *g* for 10 minutes at 4°C. Phage-containing supernatants were filtered using Acrodisc 13 mm SUPOR 0.45 μm syringe filters (Pall, 4604) into 15 mL conical tubes and supernatants frozen down as phage stocks at -80°C (100 μL filtered supernatant + 900 μL DMSO). To grow up a phage stock for plaque assays and other experiments, a pipette tip was used to scrape off a tiny portion of a frozen phage stock, which was then resuspended in 20 μL LB medium. Serial dilutions were prepared from the resuspended phage and spotted on a fresh LB top agar (LB broth Lennox base, 0.5% agar) lawn of *E. coli* EC100 in LB agar. The plate was incubated at 37°C overnight after drying at room temperature for 25 minutes. The next day a single phage plaque was picked from the top agar lawn using a P20 pipette set to 15 μL and resuspended in a 10 mL culture of exponentially growing *E. coli* EC100 at OD_{600}~0.3. The phage-added culture was incubated at 37°C with shaking overnight. The tube was spun down the next day at 15,000 x *g* for 10 minutes at 4°C. The phage-containing supernatant was filtered using an Acrodisc 13 mm SUPOR 0.45 μm syringe filter (Pall, 4604) into a 15 mL conical tube. All final phage stocks were titered on top agar lawns of *E. coli* EC100 and stored at 4°C.

To grow phage stocks of Brig1 escaper phages, single plaques formed by T4 or T6 phages on lawns of pBrig1-carrying EC100 cells were picked using a P20 pipette and resuspended in 20 μL LB medium. Serial dilutions were prepared from the resuspended phage and spotted on a fresh LB top agar lawn of *E. coli* EC100 carrying pBrig1 to maintain selection of the escaper phage. The plate was incubated at 37°C overnight after drying at room temperature for 25 minutes. The next day a single phage plaque was picked from the top agar lawn using a P20 pipette set to 15 μL and resuspended in a 10 mL culture of exponentially growing OD_{600}~0.3 *E. coli* EC100 carrying pBrig1 for continued selection. The phage-added culture was incubated at 37°C with shaking overnight and filtered the next day as described earlier to generate the escaper phage stock. Final phage stocks were titered on top agar lawns of *E. coli* EC100 and stored at 4°C. The phages used in this thesis are listed in **Table 9.5**.

9.22 Generation of T4 and T6 mutant phage stocks

T4 and T6 phage stocks were used to construct T4 Δ*a-gt*, T4 Δ*b-gt*, T4 Δ*alc* Δ*denB* Δ*gp56*, T4(C) and T6 Δ*ba-gt* mutant phage stocks. In each case, a culture of *E. coli* EC100 cells carrying a recombinant pUT18C-based plasmid was grown overnight at 37°C with shaking in 10 mL LB supplemented with 100 μg/mL carbenicillin. The pUT18C plasmid contained a cloned segment of phage T4 or T6 DNA with the desired gene deleted and ~750-1000 bp homology arms flanking the deleted genic region on either side. The overnight culture was diluted 1:50 in 10 mL

LB medium supplemented with 100 μg/mL carbenicillin. After approximately one hour of culture growth, OD_{600} was measured for the culture and confirmed to be between 0.2-0.4. The 10 mL culture was then infected with 2 μL of T4 or T6 phage stock and grown overnight at 37°C with shaking to allow wild-type phages to recombine with the plasmid. The next day, the tube was spun down at 15,000 x *g* for 10 minutes at 4°C. The phage-containing supernatant was filtered using an Acrodisc 13 mm SUPOR 0.45 μm syringe filter (Pall, 4604) into a 15 mL conical tube.

Serial dilutions of recombinant phage were prepared and spotted on a fresh top agar lawn of *E. coli* containing a pCas9 plasmid in LB agar supplemented with 25 μg/mL chloramphenicol. The pCas9 plasmid carried a type II-A CRISPR spacer targeting the phage gene that was deleted to select specifically for recombinant phage with the desired deletion. Top agar plates were incubated at 37°C overnight after drying at room temperature for 25 minutes. The next day multiple phage plaques were picked from the top agar lawn using a P20 pipette set to 15 μL and resuspended in 20 μL. 5 μL of the resuspend phage plaques were boiled in 15 μL colony lysis buffer (Pyenson et al., 2017) at 98°C for 15 minutes and then PCR checked to confirm that the desired gene was deleted, either with the deletion carried on the pUT18C recombinant plasmid or a *de novo* CRISPR-generated deletion that eliminated the appropriate gene. Serial dilutions were prepared for 1-2 correct phage plaques, which were then replaqued onto top agar lawns of pCas9 selection strains and incubated overnight at 37°C for stringent selection. The next day, a single phage

plaque was picked from the top agar lawn using a P20 pipette set to 15 μL and pipetted directly into an OD_{600} ~0.2-0.4 exponentially growing culture that maintained the same selection for the deleted phage. The phage-infected culture was grown overnight at 37°C with shaking. The next day, the tube was spun down at 15,000 x *g* for 10 minutes at 4°C. The phage-containing supernatant was filtered using an Acrodisc 13 mm SUPOR 0.45 μm syringe filter (Pall, 4604) into a 15 mL conical tube. In some cases, an arabinose inducible type I-E CRISPR-Cas expressing *E. coli* strain, ACT-01, with a pACYC184-based plasmid expressing an arabinose-inducible type I-E CRISPR spacer was used to select for the recombinant phage. In these instances, 0.2% L-arabinose was included in all media for proper phage selection. To make the T4 Δ*a-gt* phage, instead of CRISPR selection, pBrig1 was used to select for the pUT18C-recombined phage. In this case, 12.5 μg/mL chloramphenicol was used in all media. For the plasmids used to generate each mutant phage, please refer to **Table 9.3** and **Table 9.5**. All final phage stocks were titered on top agar lawns of *E. coli* EC100 and stored at 4°C. The phages used in this thesis are listed in **Table 9.5**.

9.23 Plaque assays of T-even and T-odd phages in *E. coli*

Overnight cultures were launched from single colonies in 3 mL of LB medium with appropriate antibiotic(s). Top agar lawns of *E. coli* were prepared by mixing 100 μL of overnight culture with 6 mL of LB top agar supplemented with appropriate antibiotic(s). Top agar mixtures were poured over LB agar in 10 cm plates supplemented with appropriate antibiotic(s). Where necessary, 0.2% L-arabinose

was included in the overnight media as well as in the LB top agar and the LB agar plate. Plates were dried at room temperature, partially open by a sterilizing flame, for 25 minutes for the top agar to solidify. Serial dilutions of phage stock were prepared and spotted on the top agar after drying. For imaging of plaque assays, 2.5 μL of each phage dilution was spotted on top agar using a multichannel pipette. For quantification of phage titers and isolation of single phage plaques for phage DNA sequencing, 3 μL of each phage dilution was spotted on top agar using a multichannel pipette and the plate was tilted to allow phage spots to drip down the plate for easier quantification and isolation of single plaques. In either case, plates were incubated at 37°C overnight after drying at room temperature for 25 minutes or until the plates were completely dry. Overnight plaque formation assays were imaged the next day (~16-24 hours after infection) using the FluorChem HD2 system (ProteinSimple). Plaque assay images were all auto-contrasted using Adobe Photoshop to give clearer images. In some cases, image brightness was enhanced further using Adobe Photoshop for better visualization of phage spots.

9.24 Functional selection of a T4-resistant clone in the AZ52 soil DNA library

The DNA library was generated in an earlier study using DNA extracted from an arid soil sample collected in Arizona (Brady, 2007). The library, AZ52, is comprised of large ~40 kb DNA fragments from soil microorganisms cloned into a pWEB-TNC cosmid. The insert-carrying cosmids were transformed into *E. coli* EC100 cells,

generating a soil DNA library with approximately 20 million clones, divided into megapools carrying roughly 1.25 million clones each.

Each clone within the library houses a cosmid with a soil DNA insert, which carries genes from soil-derived microorganisms. Soil-derived genes can therefore be expressed heterologously in our library system. We performed our functional screen using the coliphage T4. To grow up libraries, we scraped frozen library stocks of *E. coli* EC100 carrying megapools 3-16 of the AZ52 DNA library into separate tubes with 10 mL LB supplemented with 12.5 μg/mL chloramphenicol and grew cultures overnight at 37°C with shaking. The next day, we infected *E. coli* EC100 overnight cultures with T4 at a multiplicity of infection (MOI) of 10, high enough to kill almost all clones without *bona fide* resistance. Infections were performed in 6 mL LB top agar with 500 μL of overnight stationary culture mixed with phage at MOI 10 on LB agar plates, supplemented with 12.5 μg/mL chloramphenicol. We incubated plates overnight at 37°C and the next day we inspected surviving colonies within top agar infections. We found that only megapool 4 showed an increased number of surviving colonies upon T4 infection compared to an infection of *E. coli* EC100 cells carrying an empty pWEB-TNC cosmid (control).

Since cells may survive T4 infection due to mutations within the *E. coli* host that prevent viral infection and not due to immunity genes carried within the soil DNA cosmids, we wanted to enrich for true phage resistance genes carried on cosmids.

To eliminate false positive clones, we extracted pooled cosmid DNA from surviving colonies. To do this, we scraped top agar with surviving colonies into 50 mL conical tubes, melted the top agar in a 98°C heating block for 10-15 mins until the top agar was completely melted, and then centrifuged the tubes at ~4000 x *g* for 5 minutes at room temperature to collect cell pellets from which surviving cosmids were isolated using the QIAprep Spin Miniprep Kit (QIAGEN, 27104). Miniprepped cosmid pools were then transformed into 50 μL of electrocompetent *E. coli* EC100 cells (Lucigen) through electroporation (1 mm Bio-Rad Gene Pulser cuvette at 1.8 kV) and outgrown in 1 mL SOC medium. After 1.5 hours of outgrowth, cells were assayed for transformation efficiency by pipetting ten-fold serial dilutions of outgrowth culture on to LB agar plates supplemented with 12.5 μg/ml chloramphenicol. While plates were grown overnight at 37°C, outgrowth cultures were stored overnight at 4°C. The next day, based on the calculated transformation efficiency, transformation cultures were spread onto ten 15 cm LB agar plates supplemented with 12.5 μg/mL chloramphenicol, plating for ~30,000 colonies on each plate, for a total of ~300,000 colonies. Plates were incubated overnight at 37°C and the next day colonies from all ten plates were scraped into 20 mL LB, vortexed and inverted to mix, and then diluted to OD_{600} = 10. The OD_{600} = 10 colony mixture was then mixed 1:1 with 50% glycerol to make a -80°C freezer stock of a 1X phage-enriched DNA library for AZ52 megapool 4. This library was then grown up for re-infection with T4 and the steps described above were repeated two more times to generate a freezer stock of a 3X phage-enriched DNA library for AZ52 megapool 4.

We sampled colonies from the 3x-enriched library for phage resistance by streaking the library to single colonies on an LB agar plate supplemented with 12.5 μg/mL chloramphenicol. Sixteen single colonies were grown overnight in LB supplemented with 12.5 μg/mL chloramphenicol at 37°C with shaking. Colonies were assayed for T4 phage resistance using plaque assays (described above) with phages λ*vir* and T4. Of the sixteen colonies, twelve were found to be resistant to T4 and none to λ*vir*. Cosmids were isolated from the twelve T4-resistant clones using the QIAprep Spin Miniprep Kit (QIAGEN, 27104) and sent for Sanger sequencing by Genewiz/Azenta using the universal primers T7 and M13F40, which flank the metagenomic DNA insert within the pWEB-TNC cosmid. Sequencing the T4-resistant cosmids revealed they all contained the same metagenomic DNA insert, suggesting that they all originated from the same T4-resistant library clone.

9.25 Cosmid sequencing, assembly and gene annotation

Cosmid DNA was extracted using the QIAprep Spin Miniprep Kit (QIAGEN, 27104). DNA was sequenced using the Nextera XT DNA Library Preparation Kit (Illumina, FC-131-1024). Paired-end 2 × 75-bp sequencing was conducted using the 150-cycle MiSeq Reagent Kit v3 (Illumina, MS-102-3001) on the Illumina MiSeq platform. Geneious Prime was used to assemble the cosmid genome, using the Geneious assembler (medium sensitivity/fast) on 100,000 paired-end DNA sequencing reads. SnapGene was used to predict ORFs with ATG start codons

(minimum length: 50 amino acids) within the metagenomic DNA insert of the assembled cosmid genome. Predicted ORFs were then run through NCBI PSI-BLAST and HHpred to ascertain protein function where possible.

9.26 Subcloning of cosmid to identify T4 anti-phage system

To identify the T4-resistant gene(s) in our cosmid, we subcloned four DNA fragments (A-D) that span the entire length of the metagenomic insert sequence. DNA fragments were amplified using 10 ng of cosmid DNA as template for PCR amplification using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, F530L) with 1 M betaine and 1 µL DMSO in a 50 µL PCR reaction. Fragments were cloned into PCR-amplified pWEB-TNC cosmid backbones using NEBuilder® HiFi DNA Assembly Master Mix (NEB, E2621L). NEBuilder® HiFi DNA assembly was carried out at 50°C in a thermal cycler for 4 hours, and then 5 µL of the assembly reaction was transformed into chemically competent *E. coli* EC100 cells (Lucigen). Cells were incubated on ice for 30 minutes, heat shocked in a 42°C water bath for 30 seconds and then outgrown in SOC for 2 hours. Cells were then plated on LB agar supplemented with 12.5 µg/mL chloramphenicol and incubated overnight at 37°C. The next day, 8 colonies were picked, grown overnight in LB supplemented with 12.5 µg/mL chloramphenicol and their cosmids miniprepped the next day using the QIAprep Spin Miniprep Kit (QIAGEN, 27104). Miniprepped cosmids were sent for Sanger sequencing by Genewiz/Azenta using the universal primers T7 and M13F40, which flank the subcloned DNA fragment inserted into the pWEB-TNC cosmid

backbone. Colonies that harbored cosmids with correct insert fragments were then assayed for T4 resistance using plaque assays (see above). Plaque assays identified Fragment D as the fragment harboring T4-resistance. Fragment D was then further subdivided into Fragments D1, D2 and D3 and cloned and tested for T4 resistance as described above. Fragment D3, containing a three-gene operon, was finally identified as the minimal DNA fragment carrying T4 resistance. To determine the gene or genes responsible within the Fragment D3 operon, we generated six cosmid constructs (D3-1 to D3-6) containing different numbers and combinations of the three genes within the operon, each time being driven by the same promoter upstream of the first gene within the operon. These constructs were then tested for T4 resistance using plaque assays to identify the gene within the operon that conveyed T4 immunity. Subcloned cosmids are described in **Table 9.3**.

9.27 NCBI blastn of T4-resistant cosmid

To identify possible organisms that our metagenomic DNA comes from, we performed a nucleotide BLAST on NCBI using the algorithm for somewhat similar sequences (blastn). We performed blastn on the DNA sequences of Fragments C and D (see above).

9.28 T4 phage adsorption assay

An overnight culture of *E. coli* EC100 cells carrying pWEB-TNC or pBrig1 was diluted 1:50 in 10 mL LB medium supplemented with 12.5 μg/mL chloramphenicol. After 1 hour 15 minutes of culture growth, OD_{600} was measured for each culture and normalized to OD_{600} = 0.3. Cultures were then infected with T4 at MOI 0.01 and incubated at 37°C with shaking for 50 minutes. A 10 mL bacteria-free, media-only control (LB + 12.5 μg/mL chloramphenicol) was mixed with the same volume of T4 and incubated alongside the cultures at 37°C with shaking for 50 minutes. Phage-infected cultures were sampled at the following time points: 0-, 10-, 20-, 30-, 40- and 50-minutes post infection. At each time point, 400 μL was collected from each phage-infected culture and spun down in 1.5 mL Eppendorf tubes at 15,000 rpm for 2 minutes in a tabletop microcentrifuge at 4°C. Phage-containing supernatants were filtered using Acrodisc 13 mm SUPOR 0.45 μM syringe filters (Pall, 4604) into fresh 1.5 mL Eppendorf tubes. Filtered phage supernatants were used to prepare two sets of serial dilutions to estimate phage titers on fresh top agar lawns of *E. coli* K-12 MG1655. Top agar plates were incubated at 37°C overnight. The next day, phage plaques were counted to determine phage titers at each timepoint.

9.29 qPCR of T4 phage DNA replication

To quantify phage DNA replication within an infected *E. coli* cell, an overnight culture of *E. coli* EC100 cells carrying pWEB-TNC or pBrig1 was diluted 1:50 in 10

mL of LB medium supplemented with 12.5 μg/mL chloramphenicol. After 1 hour 15 minutes of growth, OD_{600} was measured, and the culture was normalized to OD_{600} = 0.3. 700 μL of culture was dispensed between multiple 1.5 mL Eppendorf tubes, corresponding to three replicates and multiple timepoints for each infection being monitored. These 700 μL cultures were infected with phage T4 at MOI 1 and incubated at 37°C with shaking for specified timepoints. At each timepoint, samples were removed from the incubator and tubes spun down at 15,000 rpm for 1 minute in a tabletop microcentrifuge at 4°C. Supernatants were removed and cell pellets immediately frozen down at -80°C for DNA extraction later. Additionally, 1-3 uninfected tubes for cells carrying pWEB-TNC or pBrig1 were also prepared for DNA extraction as no-phage controls for qPCR.

Total DNA was extracted from frozen *E. coli* cell pellets using the Promega Wizard® Genomic DNA Purification Kit following the protocol for Gram-negative bacteria. Extracted DNA was quantified using the Qubit™ dsDNA HS Assay Kit and each sample was normalized to 4 ng/μL. A total of 32 ng DNA was used as input for qPCR, performed using Fast SYBR Green Master Mix (Life Technologies) and QuantStudio® 3 Real-Time PCR System (Applied Biosystems) with primer pairs AA870/AA871 (T4 *gp43* target), AA872/AA873 (T4 *gp43* target) and AA387/AA388 (*E. coli* K-12 MG1655 *dxs* control) (see **Table 9.1**). For qPCR data analysis, ΔΔCt values were calculated for the two T4 qPCR targets for each replicate at each timepoint. Fold-change values were then calculated for each replicate relative to the

mean ΔΔCt value for cells carrying pWEB-TNC infected with T4 phage at the earliest timepoint post infection for a given experiment. The mean fold change of three biological replicates was plotted for each timepoint post infection.

9.30 Next-generation sequencing of phage DNA in T4-infected *E. coli* cells

Overnight cultures of *E. coli* EC100 cells carrying pWEB-TNC or pBrig1 were diluted 1:50 in 10 mL of LB medium supplemented with 12.5 μg/mL chloramphenicol. After 1 hour 15 minutes of growth, OD_{600} was measured, and cultures were normalized to OD_{600} = 0.3. Cultures were then infected at MOI 5 with T4 or T4 escaper1 for 8 minutes at 37°C with shaking, prior to centrifugation at 15,000 x *g* for 5 minutes at 4°C and subsequent freezing of cell pellets at -80°C. All cell pellets were stored at -80°C overnight until ready for genomic DNA purification using the Promega Wizard Genomic DNA Purification Kit (Promega, A1125) following the protocol for Gram-negative bacteria. Purified genomic DNA was sheared using a pre-split snap-cap 6x16 mm Covaris microTUBE (Covaris, 520045) in a Covaris S220 focused-ultrasonicator and prepared for next generation sequencing using the Illumina TruSeq Nano DNA LT kit (Illumina, 20015964). Paired-end 2 × 75-bp sequencing was conducted using the 150-cycle MiSeq Reagent Kit v3 (Illumina, MS-102-3001) on the Illumina MiSeq platform. Illumina paired-end sequencing reads were aligned to phage genomes using a custom Python script, where the recorded number of phage-

derived sequencing reads at a specific position within the phage genome was normalized to the total sequencing reads for each sample.

9.31 Phage DNA extraction for sequencing and in vitro assays

Phage genomic DNA was extracted from capsids using a previously described protocol (Jakociune and Moodley, 2018). Briefly, three tubes of 450 μL of a phage stock were first treated with DNase I and RNase A, then capsids digested with proteinase K, and phage genomic DNA was finally extracted using the DNeasy Blood & Tissue kit (QIAGEN, 69504). DNA was quantified using the Qubit™ dsDNA HS Assay Kit and assessed for quality using a nanodrop spectrophotometer.

9.32 T4 and T4 escaper1 genome sequencing and assembly

Phage genomic DNA was sequenced using the Nextera XT DNA Library Preparation Kit (Illumina, FC-131-1024). Paired-end 2 × 75-bp sequencing was conducted using the 150-cycle MiSeq Reagent Kit v3 (Illumina, MS-102-3001) on the Illumina MiSeq platform. Reads were quality-trimmed using Sickle (https://github.com/najoshi/sickle) and assembled into contigs using ABySS (https://github.com/bcgsc/abyss). Finally, contigs were mapped to a reference phage λ genome (GenBank: AF158101.6) using Medusa (http://combo.dbe.unifi.it/medusa). Automated genome annotation was performed using SnapGene and a reference phage T4 genome from NCBI (GenBank: AF158101.6). Alignment of the T4 and T4

escaper1 genomes to the reference T4 genome revealed differential mutations between the two assembled phage genomes.

9.33 Sanger sequencing of T4 and T6 escapers of Brig1 targeting

T4 or T6 phage plaques on lawns on *E. coli* EC100 cells carrying pBrig1 were isolated and resuspended in 20 μL of LB medium. Serial dilutions were prepared from the resuspended phage and spotted on a fresh LB top agar lawn of *E. coli* EC100 carrying pBrig1 to maintain selection of the escaper phage. The plate was incubated at 37°C overnight. The next day a single phage plaque was picked from the top agar lawn using a P20 pipette set to 15 μL and resuspended in 20 μL of colony lysis buffer (Pyenson et al., 2017). Resuspended phage mixtures were boiled at 98°C for 15 minutes in a thermal cycler, and 1 μL of the boiled phage mixture was then used as template for PCR amplification using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, F530L) with primers AA681/AA682 to amplify T4 *a-gt* and primers AA1115/AA1116 to amplify T6 *a-gt*. PCR products were submitted to Sanger sequencing by Genewiz/Azenta to identify mutations in *a-gt*. Wild-type T4 and T6 phage stocks were also PCR amplified at *a-gt* loci and sent for Sanger sequencing to provide reference sequences for comparison. Snapgene was used to align Sanger sequencing products of the escaper phages to wild-type *a-gt* sequences to identify escape mutations.

9.34 Brig1 structural predictions using AlphaFold2

The 261 amino acid sequence of Brig1 was run through AlphaFold2 colab (ColabFold) (Mirdita et al., 2022) using default settings (except that the amber option was turned on to improve side chain rotamers) to predict the protein structure adopted by Brig1. The highest ranked PDB structure produced by ColabFold (ptm = 0.86) was then visualized using PyMOL. Protein structure predictions of the Brig1 homologs from *Nocardiodes zhouii* and *Nocardiodes anomalus* were performed in the same way.

9.35 Purification of Brig1

Brig1 was recloned from pAM38 into the Nde1 and XhoI sites of pET21a using PCR primers that destroyed the XhoI site and added a His_6 tag immediately after the native C-terminal glycine of Brig1. The insert was verified by DNA sequencing. *E. coli* strain Rosetta (DE3) plysS was used for protein expression. Cells were grown in LB medium with 100 μg/mL ampicillin at 37°C. 0.5 mM IPTG was added to induce protein expression when $OD_{600} \approx 0.7$, followed by further growth at 37°C for 2 hours. Cell pellets were resuspended in Ni column buffer A (50mM Phosphate, 1 M NaCl, 5% Glycerol, 1 mM DTT, pH 7.5) with complete mini protease inhibitor cocktail (Roche), one tablet/1 L culture. After adding lysozyme to a final concentration of 200 mg/ml, the mixture was sonicated 3 times for 1 minute each, then centrifuged at 20,000 rpm in an SS-34 rotor for 1 hour. The supernatant was filtered and loaded

onto a Ni column (Cytiva, HisTrap HP, Cat#17524802), and eluted with a 30-minute gradient of 0 to 100% buffer B (Ni buffer A plus 500 mM imidazole, pH 7.5). Brig1-containing fractions were pooled and diluted with Heparin column buffer A (25 mM MES, 0.5 mM EDTA, 5% Glycerol, 1 mM DTT, pH 6), loaded on a Heparin column (Cytiva, HiTrap Heparin HP, Cat#17-04006-01) and eluted with a gradient from 10% to 70% Heparin buffer B (heparin column buffer A+ 2M NaCl, pH 6) over 90 minutes. The purest fractions were pooled and concentrated, then dialyzed into storage buffer (20 mM Tris, 0.5 mM EDTA, 200 mM NaCl, 20% Glycerol, 2 mM DTT, pH 8) and flash-frozen in small aliquots.

9.36 Purification of T4 alpha-glucosyltransferase

A pET15b derivative encoding N-terminally His_6 tagged bacteriophage alpha-glucosyltransferase (a-GT) was a gift from Dr. Joshua S. Chappie at Cornell University. Rosetta(DE3)plysS harboring this plasmid were grown and induced as for Brig1, but after induction grown at 20°C overnight rather than for 2 hours at 37°C. The same purification protocol as for Brig1 was followed except for a change in the pH of the heparin column buffers (A = 25 mM Hepes, 0.5 mM EDTA, 5% Glycerol, 1 mM DTT, pH 7 and B = A + 2 M NaCl, pH 7). The purest fractions were pooled and concentrated, then dialyzed into storage buffer (20 mM Tris, 0.5 mM EDTA, 200 mM NaCl, 20% Glycerol, 2 mM DTT, pH 8) and flash-frozen in small aliquots.

9.37 Purification of Brig1(Y121A, E147A) mutant

The Brig1(Y121A, E147A) mutant protein was purified according to a modified protocol. For consistency, wild-type Brig1 was purified according to this same protocol, side-by-side, and this batch of purified Brig1 protein was used only in experiments where Brig1(Y121A, E147A) was used. Both Brig1 and Brig1(Y121A, E147A) were cloned into a pET21a vector, with a His_6 tag immediately after the native C-terminal glycine of Brig1. *E. coli* strain BL21(DE3) was used for protein expression. Cells were grown in LB medium with 100 μg/mL ampicillin at 37°C overnight. The next day, a 1:100 dilution of the overnight was grown in 1 L of LB medium with 100 μg/mL ampicillin at 37°C for 3-4 hours. 0.5 mM IPTG was added to induce protein expression when $OD_{600} \approx 0.7$, followed by overnight growth (~16 hours) at 18°C. Cells were pelleted at 4500 rpm at 4°C (Eppendorf Centrifuge 5810 R) for 15 minutes. Cell pellets were resuspended in 20 mL of lysis buffer (50 mM HEPES, pH 7.7, 150 mM NaCl, 10% glycerol, 1 mM TCEP, 30 mM imidazole, 2 Roche mini protease inhibitor tabs EDTA free, 0.5 mg/mL lysozyme) and incubated on ice for 1 hour with shaking. The resuspended pellets were then sonicated using a Qsonica Q500 sonicator (70% amplitude with 10 seconds on, 30 seconds off for 2.5 minutes). The sonicated samples were spun down on at 12,000 rpm at 4°C (Eppendorf Centrifuge 5810 R) for 30 minutes and the supernatant run through a gravity column loaded with 3 mL of HisPur Ni-NTA Resin (Thermo Scientific, Cat# 88222). Before passing supernatant, the column was equilibrated with equilibration buffer (50 mM HEPES, pH 7.7, 150 mM NaCl, 10% glycerol, 1 mM TCEP, 30 mM

imidazole). Then, the ~20 mL of sonicated cell pellet supernatant was passed through the column. The column was washed twice with 25 mL wash buffer (50 mM HEPES, pH 7.7, 500 mM NaCl, 10% glycerol, 1 mM TCEP, 30 mM imidazole) and then eluted with 20 mL elution buffer (50 mM HEPES, pH 7.7, 150 mM NaCl, 10% glycerol, 1 mM TCEP, 300 mM imidazole). The eluted protein was concentrated to < 500 μL using an Amicon Ultra-4 Centrifugal Filter, 10 kDa MWCO (Millipore, Cat# UFC801024), with multiple rounds of centrifugation at 4300 x *g* for 10 minutes at 4°C (Eppendorf Centrifuge 5810 R). The concentrated eluant was run on an ÄKTA pure™ chromatography system (Cytiva) fitted with a Superdex™ 75 Increase 10/300 GL column (Cytiva, Cat# 29148721) using storage buffer (50 mM HEPES, pH 7.7, 150 mM NaCl, 10% glycerol, 1mM TCEP). Two peaks, corresponding to fractions 17-20 and 22-27, were collected and separately pooled. Pooled fractions were concentrated to < 500 μL using an Amicon Ultra-0.5 Centrifugal Filter, 10 kDa MWCO (Millipore, Cat# UFC501096), with multiple rounds of centrifugation at 13,000 x *g* for 5 minutes in a tabletop microcentrifuge at 4°C. For both Brig1 and Brig1(Y121A, E147A), the second peak was determined to be free of nucleic acid contamination via nanodrop and found to contain pure protein (~29 kDa) by a Coomassie gel. Concentrated protein was flash-frozen in small aliquots and stored at -80°C for future use.

9.38 Annealing of ssDNA oligonucleotides

To generate dsDNA substrates for MfeI digestion and for DNA glycosylase assays, complementary ssDNA oligonucleotides were annealed. Briefly, 1:1 molar ratios of top and bottom oligonucleotides (25-50 μM each) were mixed in a 60 μL reaction containing NaCl to a final concentration of 100 mM. The reaction was heated at 80°C for 20 minutes in a water bath or thermal cycler and then allowed to cool very slowly to room temperature. Annealed oligonucleotides to be used in DNA glycosylase assays were purified using an oligonucleotide cleanup kit (Zymo Research, Oligo Clean & Concentrator Kit, Cat# 11-380) according to the manufacturer's instructions.

9.39 Generation of glucosylated ssDNA and dsDNA oligonucleotides

We tested the activity of alpha-glucosyltransferase (a-GT) on both single- and double-stranded DNA as previous studies only tested dsDNA substrates (Dai et al., 2013). The ssDNA stranded substrates were hmdC_18, hmdC_60_MfeI and hmdC_60_MfeI_Bot, which are 18mer and 60mer oligonucleotides, each containing a single hmC residue (**Table 9.1**). The dsDNA substrates were hmdC_60_MfeI annealed to Bot_MfeI_60 and hmdC_60_MfeI annealed to hmdC_60_MfeI_Bot (**Table 9.1**). Substrate DNAs (100 μM for ssDNA and 50 μM for dsDNA) were mixed at a 1:1 molar ratio with a-GT in 1X NEBuffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9) supplemented with 2

mM UDP-Glucose (NEB, supplied with NEB T4 beta-glucosyltransferase (b-GT)). All samples were incubated at 37°C overnight, then purified with an oligonucleotide cleanup kit (Zymo Research, Oligo Clean & Concentrator Kit, Cat# 11-380) according to the manufacturer's instructions. A subset of the ss- and dsDNA substrates were also treated with b-GT (NEB, Cat# M0357S) following the supplier's instructions and purified as described above.

Modification by a-GT (or b-GT) was monitored by digestion with MfeI-HF (NEB, Cat# R3589S), which is blocked by the presence of glucosylated hmC but not by hmC (the modified C in hmdC_60_MfeI and in hmdC_60_MfeI_Bot is within an MfeI site). Before digestion, single-stranded a-GT- or b-GT- treated samples were annealed to Bot_MfeI_60 or to hmdC_60_MfeI_Bot. Approximately 1.5 μg of each sample was digested with MfeI-HF for 1 hour, then electrophoresed on a 10% TBE gel (Invitrogen, Cat# EC6275BOX) at 140 V for 35 minutes. Gels were stained with 2 μg/mL ethidium bromide for 20 minutes, extensively rinsed with distilled water (3X for 10 minutes each), and then scanned using a ChemiDoc MP imager (BioRad) set to UV trans illumination and the machine's 605/50 filter to detect ethidium bromide or using the Amersham ImageQuant 800 set to UV fluorescence.

9.40 DNA glycosylase assays with ssDNA oligonucleotides: detection of the abasic site with an aldehyde-reactive probe

We used an aldehyde-reactive fluorescent probe, AZDye 488 Hydroxylamine, (fluoroprobes.com) to detect removal of a base from the phosphodiester backbone in the absence of DNA cleavage. The dye was dissolved in distilled water to form a 10μg/μL stock solution. DNA glycosylase reactions were carried out in a reaction buffer containing 25 mM HEPES, pH 7.5, 0.4 mM EDTA, 2% glycerol, 1 mM DTT and 50 mM KCl, in a total reaction volume of 50 μL. The final DNA concentrations were 2 μM. Brig1 was added to single-stranded a-GT- and b-GT-treated hmdC_60_MfeI to a final concentration of 35 μM, while 2 μL (10 units; 5 units/μL) of hSMUG1 (NEB, Cat# M0336S) was added to dU_60 as a positive control. Reactions were incubated overnight at 37°C, after which 2 μL AZDye 488 dye was added, followed by incubation at 37°C for 30 minutes. 1/10 volume of 10% SDS was then added and incubated for another 30 minutes and purified by phenol/chloroform extraction. Samples were then treated with an oligonucleotide cleanup kit (Zymo Research, Oligo Clean & Concentrator Kit, Cat# 11-380) according to the manufacturer's instructions, eluted with 15 μL nuclease-free water, mixed with loading dye and electrophoresed for 45 minutes at 180 V on a 10% TBE gel (Invitrogen, Cat# EC6275BOX). The gel was stained with 2 μg/mL ethidium bromide for 20 minutes, extensively rinsed with distilled water (3X for 10 minutes each), then scanned using a ChemiDoc MP imager (BioRad) set to UV trans illumination and the machine's 605/50 filter to detect ethidium bromide and then using blue epi illumination with the

530/28 filter for the AZDye 488 fluorescent probe. Oligonucleotides used in this assay are listed in **Table 9.1**.

9.41 DNA glycosylase assays with ssDNA and dsDNA oligonucleotides: detection by NaOH- or Endonuclease IV-mediated cleavage of the abasic site

DNA glycosylase reactions were carried out in a reaction buffer containing 45 mM HEPES, pH 7.5, 0.4 mM EDTA, 2% glycerol, 1 mM DTT and 50 mM KCl, in a total reaction volume of 50 μL. The final ssDNA or dsDNA concentrations were 1 μM. Brig1 or Brig1(Y121A, E147A) was added to a final concentration of 1 μM, while 1 μL (5 units) of hSMUG1 (NEB, Cat# M0336S) was added as a positive control. Reactions were incubated overnight at 37°C, unless stated otherwise (30 mins or 2 hours). Following enzymatic incubation, one set of samples was directly processed with an oligonucleotide cleanup kit (Zymo Research, Oligo Clean & Concentrator Kit, Cat# 11-380) according to the manufacturer's instructions. A second matched set of samples was treated with NaOH before cleanup: 25 μL of 0.5 M NaOH was added to each 50 μL sample and then heated at 90°C for 30 minutes before purification with the oligonucleotide cleanup kit (Zymo Research, Oligo Clean & Concentrator Kit, Cat# 11-380) according to the manufacturer's instructions. All samples were eluted from the cleanup columns in 15 μL nuclease-free water. 5 μL of each was mixed with loading dye and loaded onto a 10% TBE gel (Invitrogen, Cat# EC6275BOX) and electrophoresed at 140 V for 35 minutes. Gels were stained with 2 μg/mL ethidium bromide for 20 minutes, extensively rinsed with distilled water (3X for 10 minutes

each), and then scanned using a ChemiDoc MP imager (BioRad) set to UV trans illumination and the machine's 605/50 filter to detect ethidium bromide or using the Amersham ImageQuant 800 set to UV fluorescence. For Urea-PAGE gels, eluted samples were first denatured by mixing 5 μL of purified sample with 5 μL of 2X TBE Urea Sample Buffer (Invitrogen, Cat# LC6876) and then heated at 70°C for 3 minutes. Denatured samples were loaded onto a 6% TBE-Urea gel (Invitrogen, Cat# EC6865BOX) and electrophoresed at 140 V for 35 minutes. Gels were soaked in ethidium bromide and rinsed with distilled water as described above, before imaging with the Amersham ImageQuant 800 set to UV fluorescence. For all gels, DNA ladders were made by mixing 20-, 40- and 60-bp ssDNA or dsDNA oligonucleotides and loading them onto their corresponding gels at ~100 ng each oligonucleotide per load. Oligonucleotides used in these assays are listed in **Table 9.1**.

For abasic site detection by NEB Endonuclease IV (Endo IV), reactions were set up as described above and incubated with Brig1 overnight. Three matched sets of reactions were set up. After overnight incubation, one matched set of samples was treated with NaOH as described above and purified using the Zymo Research Oligo Clean & Concentrator Kit (Cat# 11-380) according to the manufacturer's instructions. The remaining two matched sets of samples were processed directly using the oligonucleotide cleanup kit. The purified samples were then incubated at 37°C for 4 hours in a 50 μL reaction with 1X NEBuffer 3 (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9), with or without 50 units of NEB Endo IV (5 μL; 10

units/μL; Cat# M0304S). After 4 hours, reactions were purified using the Zymo Research Oligo Clean & Concentrator Kit according to the manufacturer's instructions. All the purified samples were then loaded onto a 10% TBE gel (Invitrogen, Cat# EC6275BOX), electrophoresed at 140 V for 35 minutes, stained with ethidium bromide as described above and imaged with the Amersham ImageQuant 800 set to UV fluorescence.

9.42 High resolution mass spectrometry of hSMUG1- and Brig1-treated ssDNA oligonucleotides

DNA glycosylase reactions were carried out in a reaction buffer containing 45 mM HEPES, pH 7.5, 0.4 mM EDTA, 2% glycerol, 1 mM DTT and 50 mM KCl, in a total reaction volume of 50 μL. Reactions were performed with 18mer ssDNA oligonucleotides: dU_18, hmdC_18 and a-GT treated hmdC_18 (**Table 9.1**). The final ssDNA concentration in each reaction was 2 μM. Brig1 was added to a final concentration of 2 μM, while 2 μL (10 units) of hSMUG1 (NEB, Cat# M0336S) was added as a positive control. A no-enzyme reaction was used as a negative control. 2 x 50 μL reactions were set up for each reaction condition with dU_18, while 8 x 50 μL reactions were set up for each reaction condition with hmdC_18 and a-GT treated hmdC_18. Reactions were incubated overnight at 37°C. After overnight incubation, all matched samples were pooled and processed with an oligonucleotide cleanup kit (Zymo Research, Oligo Clean & Concentrator Kit, Cat# 11-380) according to the manufacturer's instructions.

For mass spectrometry, purified oligonucleotide samples were dried using vacuum centrifugation and dissolved in 50/50 water/acetonitrile with 0.001% triethylammonium bicarbonate. The pH of the solution was found to be comparable to that of deionized water. The samples were introduced to the mass spectrometer by manual injection using a Hamilton syringe applying pressure by hand at approximately 10 μL/min. Samples were analyzed using an orbitrap Ascend tribrid mass spectrometer (Thermo Scientific) operating in negative mode. Spectra were recorded in the mass range 600-1300 m/z at 120,000 resolution. A blank injection was introduced after each sample to eliminate carryover.

Raw data was inspected using the Xcalibur Quality Browser (Thermo Scientific) and spectra were summed as necessary to provide representative spectra with a sufficient signal-to-noise ratio (S/N). Spectra were further processed using UniDec deconvolution software (Marty et al., 2015) with the following parameters: sampling resolution and peak FWHM were both set to 0.1, adduct mass was defined as -1.007276 Da, and charge states were defined 4-12 based on observations from the raw data. The m/z range was adjusted to fit the data and to exclude singly charged noise. Apart from the mass of the oligonucleotides, additional masses from metal adducts were also observed.

9.43 Generation of uracil-containing PCR products

PCR products containing deoxyuridine were generated for use in DNA glycosylase assays using NEB Q5U® Hot Start High-Fidelity DNA Polymerase (Cat# M0515S). PCR was performed using 10 ng of purified T4 genomic DNA as the template, using primers AA1122 and AA1123 which amplifies the T4 *gp24* gene to give a 1324 bp PCR product. PCR was performed following the manufacturer's guidelines with 2 mM each of dATP, dGTP and dCTP and varying μM ratios of dUTP:dTTP (0:200, 100:100, 125:75, 150:50, 175:25 and 200:0 μM). PCR products were run on a 1% agarose gel at 100 V for 45 minutes and then gel purified using the QIAquick Gel Extraction Kit (Qiagen, Cat# 28704).

9.44 DNA glycosylase assays with phage DNA, cosmid DNA or gel-purified PCR products

All reactions were performed in 50 μL reaction volumes in a reaction buffer containing 45 mM HEPES, pH 7.5, 0.4 mM EDTA, 2% glycerol, 1 mM DTT and 50 mM KCl. Assays were performed by incubating 50-500 ng of extracted phage genomic DNA from capsids, miniprepped pWEB-TNC cosmid DNA or gel purified PCR products with different concentrations (2-1600 nM) of purified Brig1 or Brig1(Y121A, E147A) or with 10 units of NEB hSMUG1 (control). Reactions were incubated in a thermal cycler at 37°C for 30 minutes or at 37°C for 30 minutes plus an additional 20 minutes at 65°C to cleave DNA at abasic sites. Reactions were then mixed with 10 μL of purple 6X loading dye with no SDS (NEB #B7025S) and the

entire reaction volume was loaded onto a 1% agarose gel containing ethidium bromide. The gel was run for 70 mins at 85 V and then imaged using a UV gel imager (Amersham ImageQuant 800 set to UV fluorescence). Where SDS was used for protein denaturation, all steps were carried out as described above except, before gel loading, a purple 6X gel loading dye containing a final 1X concentration of 0.08% SDS (NEB #B7024S) was used instead of loading dye without SDS.

9.45 Brig1 multiple sequence alignment and phylogenetic tree construction

Brig1 homologs were obtained using the NCBI PSI-BLAST protein homology search. Homologs were then subjected to a multiple sequence alignment (MSA) using MUSCLE v5 with 16 maximum iterations via the Geneious Prime software. A tree was built with the alignment output file via IQ-TREE 1.6.12 (Trifinopoulos et al., 2016) using the LG4M model with 1000 bootstrap alignments. The online tool ITOL (Letunic and Bork, 2021) was used for visualization of the resulting tree.

9.46 Brig1 neighborhood analysis

Gene neighborhoods of the Brig1 homologs from above (10 genes upstream and 10 genes downstream of each homolog) were constructed using a custom Python script. In brief, the script parses a blastp result XML file for accession numbers of each of the hits. For each hit accession, the script obtains the corresponding nucleotide accession from which the protein accession is derived.

Finally, all annotated features within the nucleotide accession that are labelled as ‘CDS’ or ‘tRNA’ are built into a list, including their position within the nucleotide entry and their feature name. From this list, neighbors of the initial protein hit (10 genes upstream and 10 genes downstream) are extracted and built into a TSV file for subsequent analysis.

Table 9.1 Oligonucleotides used in this study

Oligonucleotide	Sequence
AA66	ccggagcgtagcgaccgagtgagctagctatgatcggcacgtaagaggttc
AA67	ggcccctcggcttgaacgaattgttagacattacgccccgccctgccac
AA68	tagctagctcactcggtcgctac
AA69	tgtctaacaattcgttcaagccgag
AA77	accgcacgccagcagcgcctgctgcccctgctctcc
AA78	actcggagagcaggggcagcaggcgctgctggcgtg
AA87	cggaaatcactcccgggtatatg
AA88	gagtcagttgcatcagtcacaagg
AA107	AAACtacggcgtaattccgcatcagtaagcgcatG
AA108	AAAACatgcgcttactgatgcggaattacgccgta
AA132	AAACagggatgcaccattctgagatgtttttattG
AA133	AAAACaataaaaacatctcagaatggtgcatccct
AA138	caatggacagaatcaccgattctc
AA139	tcatagatccaccccgtaaatc
AA145	agcgaattcgagtatttcaggagttcagccatgagtactgcactcgcaacg
AA146	gtcgactctagaggatccccgggtaccgagtcatgctgccaccttctgctc
AA147	agcgaattcgagtatttcaggagttcagccatgacaccggacattatcctg
AA148	gtcgactctagaggatccccgggtaccgagtcatcgccattgctccccaaatac
AA149	ggctgaactcctgaaatactcg
AA150	ctcggtacccggggatcctctag
AA153	AAACactcgtcagaatgaatattatcaagcagcaG
AA154	AAAACtgctgcttgataatattcattctgacgagt
AA164	AAAACcagaaaggtcgttttctggctggtcagagg
AA207	agcgaattcgagtatttcaggagttcagccatgaacgcttattacattcagg
AA219	cagtcgggaaacctgtcgtg
AA220	gcatagttaagccagccccgac
AA246	caacacccgctgacgagcttag
AA266	CAAAagatgcagactggactgcgacc
AA267	aCCCAagatgcagactggactgcgacc
AA268	atGGGAagatgcagactggactgcgacc
AA269	cagTTTAagatgcagactggactgcgacc

AA270	cgtaTTACagatgcagactggactgcgacc
AA271	acttacAGCCagatgcagactggactgcgacc
AA272	CCGCagatgcagactggactgcgacc
AA273	tGATCagatgcagactggactgcgacc
AA274	caCCAGagatgcagactggactgcgacc
AA275	tgaAACGagatgcagactggactgcgacc
AA276	acgaCTGGagatgcagactggactgcgacc
AA277	atcgtGGTGagatgcagactggactgcgacc
AA278	AGATagatgcagactggactgcgacc
AA279	aCTCTagatgcagactggactgcgacc
AA280	caTCGAagatgcagactggactgcgacc
AA281	atcTATAagatgcagactggactgcgacc
AA282	CAAAttcattaaaccacgccagcagc
AA283	aCCCAttcattaaaccacgccagcagc
AA284	gtGGGAttcattaaaccacgccagcagc
AA285	tgcTTTAttcattaaaccacgccagcagc
AA286	tgcaTTACttcattaaaccacgccagcagc
AA287	tacgtAGCCttcattaaaccacgccagcagc
AA288	CCGCttcattaaaccacgccagcagc
AA289	aGATCttcattaaaccacgccagcagc
AA290	atCCAGttcattaaaccacgccagcagc
AA291	tacAACGttcattaaaccacgccagcagc
AA292	tgcaCTGGttcattaaaccacgccagcagc
AA293	tcagcGGTGttcattaaaccacgccagcagc
AA294	AGATttcattaaaccacgccagcagc
AA295	gCTCTttcattaaaccacgccagcagc
AA296	caTCGAttcattaaaccacgccagcagc
AA297	tcaTATAttcattaaaccacgccagcagc
AA318	AAACagataccagcttcacgctggcgtggatgccG
AA319	AAAACggcatccacgccagcgtgaagctggtatct
AA320	AAACcaggggtgttaccactaccgcaggaaaaggG
AA321	AAAACccttttcctgcggtagtggtaacacccctg
AA324	cgcagcaaactcaccattac
AA325	atggagtcccggatttatcc
AA326	ggtatcgccttcattaaaccac
AA327	cgcatcctcacgataatatccgg
AA331	ggccttcctcgtgatgaag

AA332	cactttctggctggatgatgagg
AA333	cgcctacccggatattatcgtgaggatgcgttttataagcgtcgactg tttcctg
AA334	TAGGTGACGTCTCTCGTCAGGTTG
AA336	tcaatttttcgtaatagcgcatctc
AA338	gaatcagatatcttgctgaactgtcag
AA341	ccgggcagtttgttgcatg
AA342	cctggtagcaaaccggtaatacac
AA349	aaacCACCATCAGTTCAAGACGACGCAGCACCTCg
AA350	aaaacGAGGTGCTGCGTCGTCTTGAACTGATGGTG
AA383	AAACactcgtcagaatgaatattttcaagcagcaG
AA384	AAAACtgctgcttgaaaatattcattctgacgagt
AA385	CGCTTATGCTGGAAAGAAGC
AA386	TGACGTTTCTAATCGGAAGC
AA387	cttcatcaagcggtttcaca
AA388	cgagaaactggcgatcctta
AA396	tatcattctacatttaggcgctg
AA397	agcacactgagacttgttgagttc
AA398	aagatggcagcgcctaaatgtagaatgatagacagatccagtcgcgct g
AA399	aacatggaactcaacaagtctcagtgtgctatgatttaaatggtcagt attgagcc
AA432	CATGCGCCTTCTCCCTGTACC
AA433	TAACGTGTGACCGCATTCAAAATG
AA549	CATCTGCTTCTGCTTTCGCCAC
AA550	CGCCTTATCGACATACTTAATCAGCC
AA553	GACAGGAAGAACTTGCCGC
AA554	CTCTTTCCATGCCGCTTCAAC
AA555	CATTGAGAGTGAGCTGGATACG
AA556	GTAGGTACTCAGTCCGGCTTC
AA557	GATAGGCTCCGGTCGTATGG
AA558	TCCGTCAGGCTGCTGATC
AA559	ccctatagtgagtcgtattatgcggc
AA560	gtcgtgactgggaaaaccctggc
AA614	CGCGGCCGCATAATACGAC
AA615	ATAGGCGTATCACGAGGCCC
AA625	cgcggccgcataatacgactcactatagggGTTCATCCCTTGCCATTC GCC

AA627	cgcggccgcataatacgactcactatagggAAGTGGCCCTCGACGTAGGAG
AA629	cgcggccgccagggttttcccagtcacgacAGGGACTACTGGGCTGAG
AA630	AGCTTATCGATGATAAGCGGTCAAAC
AA640	cgcggccgcataatacgactcactatagggCGAGTGCAATGTCGACTCTTTC
AA643	cgcggccgccagggttttcccagtcacgacTTCTCCGCGCAAAGTGCGAC
AA644	cgcggccgcataatacgactcactatagggTTCTGGGCCGAGGTCAGAC
AA646	cgcggccgccagggttttcccagtcacgacAGGTTGTCTGCGTCGAGCAG
AA672	cgcggccgcataatacgactcactatagggTTCACAGCTGAGTCCCCCTC
AA673	CCTTCACCGGCCACCTCGATCGGGCCGAGCTCCACGCGACCGTCACCAACGTC
AA674	GCTCGGCCCGATCGAGGTGGC
AA675	CTTCACCGGCCACCTCGATCGGGCCGAGCGAGCACGGCGTGAAAGTG
AA676	cgcggccgcataatacgactcactatagggTCACCTCCTGGTGGAGC
AA677	CAGACAGTCACCTCCTGGTG
AA678	GCACCCGCTCCACCAGGAGGTGACTGTCTGATGAGTGCACGCGAACGG
AA681	GCATATCTGGAAAAAGCAAAATTG
AA682	AAGGGCCGAAGCCCTTTATTTTG
AA693	cgaattcgctagcccaaaaaaacg
AA694	gcggataacaatttcacacaggaaacagctgttgagaggtgagcaatgcgtaaaatc
AA695	gggtgttggcgggtgtcggggctggcttaatcataatcccagcaccagttgtc
AA696	agctgtttcctgtgtgaaattgttatcc
AA697	ttaagccagccccgacacc
AA816	tttttgggctagcgaattcgATGAGTGCACGCGAACGG
AA817	gaggatccccgggtaccgagTCAGCCCAGTAGTCCCTG
AA870	CAGCTCGACGATAAGTTAGC
AA871	AGGTGAAGAGTCTGTCCAAG
AA872	TACGTACCGATGCTAACTGG
AA873	CCTCCAATATCTTGGAGAAC
AA893	cggggctggcttaactatgcGCACTGCGTTTACGTTCATATTTTC
AA894	AACTACAACGGATGGTATCTTTCC
AA895	AGATACCATCCGTTGTAGTTCGTTTTCTGCATTTAAACTTTCC
AA896	cacgacaggtttcccgactgAGTAGTTAAAATGCCACGTACTCG
AA897	cggggctggcttaactatgcTTGCGCCCTTGAAGTTCCTTC

AA898	GCGGAAGAAATCTTTAAACTTTATTATC
AA899	AGTTTAAAGATTTCTTCCGCAAAGTGAGCCATTAGTTTTCCTTTC
AA900	cacgacaggtttcccgactgTAAGCGTCTTATTAATGTATCATGGG
AA901	atttcacacaggaaacagctATGCGTATTTGCATTTTTATGGCTC
AA902	aagctcgtcagcgggtgttgTTATTTTGTAATAATGTCAAACTGTTC
AA903	agctgtttcctgtgtgaaattgttatc
AA904	atttcacacaggaaacagctATGAAAATTGCTATAATTAATATGGGTA ATAATG
AA905	aagctcgtcagcgggtgttgTTATAAATCAATAGCTTTTTTGAACGC
AA906	GTTGATCATTGGTGGAAAACTGG
AA907	cggggctggcttaactatgcCATCCAAAGTTTTCTTGGTAGGC
AA908	GTTTTCCACCAATGATCAACGTCACTAATCATTTAAACCTCAATTG
AA909	cacgacaggtttcccgactgTGAAAACTACTGCGCTGAGC
AA922	AGCTTCTGCTTTAAAGAACAGTTTG
AA923	cggggctggcttaactatgcCTCAAGAATTTCACCAGTATTTTCTTC
AA924	TGTTCTTTAAAGCAGAAGCTAATGCAAATACGCATAGTTTTCCTC
AA925	cacgacaggtttcccgactgGGAAAGCATGAAAGAGTTCTGG
AA932	GCAAAGAAAGCCGAATGGCAAG
AA933	cggggctggcttaactatgcAGTTTCTTCTGACTGCTTTTTGCC
AA934	TGCCATTCGGCTTTCTTTGCACCCATATTAATTATAGCAATTTTC
AA935	cacgacaggtttcccgactgTTGTTCGTTAAAGATATTCCGGG
AA940	aaacCAGCCTTAGCAATGTCTAAATTTATACCTTg
AA941	aaaacAAGGTATAAATTTAGACATTGCTAAGGCTG
AA954	AGGGTAGAGCGCGAAGTTCAGAATGTCCCG
AA955	TCTGAACTTCGCGCTCTACCCTTGGCACTC
AA956	ACCCGCACCCAGGCGGCGCTCGGCCTCTTGTG
AA957	AGCGCCGCCTGGGTGCGGGTCGAGCACTTTG
AA958	AGGGTAGAGCTGGAAGTTCAGAATGTCCCG
AA959	TCTGAACTTCCAGCTCTACCCTTGGCACTC
AA960	GGTAGAGCTCGAAGGCCAGAATGTCCCGATGGCCGGCGGATGTGTC
AA961	TCTGGCCTTCGAGCTCTACCCTTGGCACTC
AA962	cacgacaggtttcccgactgGGTCATGTACATATCGTTTTACATC
AA963	TTATGCATGTAAGTAATTTTACAGC
AA964	AAAATTACTTACATGCATAAAAGTTGTAAATCCATCATAAAGTCCTC
AA965	cggggctggcttaactatgcCGGGTGAATACATTTCATACGATG
AA974	accgGCGATGTTTTCCCGTCAAAGGATTAACGATGA
AA975	actcTCATCGTTAATCCTTTGACGGGAAAACATCGC

AA976	accgTTGATTTTCTAAAAAGACTTTCATCTCTTCAA
AA977	actcTTGAAGAGATGAAAGTCTTTTTAGAAAATCAA
AA978	accgAATATCGAAATCGTCGTCTTGAGCAAGCCCGT
AA979	actcACGGGCTTGCTCAAGACGACGATTTCGATATT
AA988	caggagctaaggaagctaaaatgagggaagcggtgatcg
AA989	caataactgccttaaaaaaattatttgccgactaccttggtg
AA990	ttttttaaggcagttattggtgcc
AA991	tttagcttccttagctcctgaaaatc
AA992	tttttgggctagcgaattcgatgaaatacattggagcgcacg
AA993	gaggatccccgggtaccgagtcaggctaccgcttttcagtttg
AA994	tttttgggctagcgaattcgatgaaatttgtctcttttaatatcaacg g
AA995	gaggatccccgggtaccgagttagcggcggaaggtcgc
AA1002	accgCCCGCAGTATTTTGGATGAGCCAATACATCCC
AA1003	actcGGGATGTATTGGCTCATCCAAAATACTGCGGG
AA1004	accgACACATCAAACTGACTCACGTTGGATTTCGGT
AA1005	actcACCGAAATCCAACGTGAGTCAGTTTGATGTGT
AA1020	atgtatatctccttcttaactctagagg
AA1021	taatggctgttttggcggatg
AA1068	aaacAGTTTAAAGATTTCTTCCGCATTCATTCCAg
AA1069	aaaacTGGAATGAATGCGGAAGAAATCTTTAAACT
AA1088	aaacTAAAACAGGTGTTAAAACTATTGAAATTATg
AA1089	aaaacATAATTTCAATAGTTTTAACACCTGTTTTA
AA1090	aaacGTATTCATCATTAAGAGCGCCAAAAATAAAg
AA1091	aaaacTTTATTTTTGGCGCTCTTAATGATGAATAC
AA1104	tttttgggctagcgaattcgATGATTAGTGACTCTATGACAGTTGAAG
AA1105	gaggatccccgggtaccgagTTAAGCGTATTTTCCTACATAATC
AA1115	ggagcttcggctcctatattg
AA1116	gcaagaaaaagtatcacatatcc
AA1171	cggggctggcttaactatgcgaatcatctcgccagtttc
AA1172	catccaatagcagttgattaatgcc
AA1173	taatcaactgctattggatgagtcggaaacatatcaagggc
AA1174	cacgacaggtttcccgactggtcttcatttggggcttgc
AA1196	AAACtatgtatgcatttctcaatttaattaatcaG
AA1197	AAAACtgattaattaaattgagaaatgcatacata
AA1198	AAACgacctcggaaaacttacgacagaacaaattG
AA1199	AAAACaatttgttctgtcgtaagttttccgaggtc

AA1320	caccaccaccaccaccac
AA1321	atgtatatctccttcttaaagttaaac
AA1331	tttgtttaactttaagaaggagatatacatATGAGTGCACGCGAACGG
AA1332	agccggatctcagtggtggtggtggtggtgGCCCAGTAGTCCCTGCAG
JW488	ctgcagacaagcccggccgg
JW489	attaagctagcactgtacctaggactg
JW1059	agctgtttcctgtgtgaaattgttatc
JW1187	gctcagtcctaggtacagtgctagcttaatCTAGCTGAGACAAATAGT GCGATTAC
JW1188	tcgccttaggccggccgggcttgtctgcagCTTTCTCAAGTTATCATC GGCAATG
JW1191	aaacTGCTTGCTGAGGTTTGCACCGGTGTGGCTCg
JW1192	aaaacGAGCCACACCGGTGCAAACCTCAGCAAGCA
JW1209	caacacccgctgacgagctta
JW1370	AAACagaggcagaactggcagacgacatggaaaaG
JW1371	AAAACttttccatgtcgtctgccagttctgcctct
JW1377	ACCGGATATCCCACAGGTGAGC
JW1401	aaacGCGCTAACTGCGGTCAGAAGCTGCATGTGCg
JW1402	aaaacGCACATGCAGCTTCTGACCGCAGTTAGCGC
JW1406	tggcgggtgtcggggctggcttaactatgcGTTGTCGAATCCAATCGT ATCCAG
JW1409	cagtcgggaaacctgtcgtg
JW1410	gcatagttaagccagccccgac
JW1413	atgcagctggcacgacaggtttcccgactgATACAGGCATTCGGTACA GAGCG
JW1417	TGGTCGATGCTGAGCTGGTGGACACGCGCTGGCTGGTGGACGTTAAAG CTCGCTCGACGC
JW1418	CCAGCGCGTGTCCACCAGCTCAGCATCGACCACCAGCAAGTTCACGTG TGTGAGCACTGC
JW1480	gcggataacaatttcacacaggaaacagctatgaaaaacatcgccgca cag
JW1481	gggctttactaagctcgtcagcgggtgttgttaaacggaagcaccctt caatccgaacttagctttgatttctgc
JW1484	ctcgaattcgctagcccaaaaaaac
JW1485	ctcggtacccggggatcctctag
JW1491	ccgaccaaatcaaccttactg
JW1518	ctggatttacggggtggatctatgaTGTAGGCTGGAGCTGCTTC
JW1519	gcccgtcatacacttgctcctttcaTGGTCCATATGAATATCCTCCTT AG
JW1520	caaaactcgacctgacaaacacagactggatttacggggtggatc
JW1521	ccaggtaatgaataattgcctctttgcccgtcatacacttgctcc

JW1522	gcggataacaatttcacacaggaaacagctATGAGTGAtgtcgccgagacac
JW1523	gggctttactaagctcgtcagcgggtgttgTTACGCCTCCTCCAGGGTCATAC
JW1524	GAGCAGGTAATAACGCCCTTCG
JW1525	TTCCGCCACGAAGGGCGTTATTACCTGCTCGCATATAAATCCAACTGGTTGGGTGAAGAC
JW1546	AAACgcggccttttacacatgaccttcgtgaaagG
JW1547	AAAACctttcacgaaggtcatgtgtaaaaggccgc
JW1548	atgcagctggcacgacaggtttcccgactgtgtatgacgctctggtggtg
JW1549	tttcacgaaggtcatgtgtaaaag
JW1550	tgcggccttttacacatgaccttcgtgaaaGCTGGTGGcaggaggtcgcgctaacaac
JW1551	tggcgggtgtcggggctggcttaactatgccgagggtgatcggagtaatcag
JW1552	AAACcctttcctgataagcagaatggcatcgttcG
JW1553	AAAACgaacgatgccattctgcttatcaggaaagg
JW1554	atgcagctggcacgacaggtttcccgactggcgatcagcagagtgttaatc
JW1555	tggcgggtgtcggggctggcttaactatgcacatgctctgcttatagcaatttc
JW1556	atgcagctggcacgacaggtttcccgactggcacaaacctgattccaatttgag
JW1557	tggcgggtgtcggggctggcttaactatgcgcatcatcaatgaaaaccagcag
JW1558	aaacGATTATGTACCGAGGAAGAATGTCGCTGGAg
JW1559	aaaacTCCAGCGACATTCTTCCTCGGTACATAATC
JW1561	GCCAGCGCGTGTCCACCAGCTCAGCATCGACCACCAGCTCCTTTGGCGTTTCCCGATGTC
JW1562	GATGCTGAGCTGGTGGACACGCGCTGGCTGGTGGTATCGCGAAAATGTATTCAGAAAATG
JW1567	tacccgtttttttgggctagcgaattcgagtatttcaggagttcagccatgaac
JW1568	gtcgactctagaggatccccgggtaccgagcgttttatacctctgaatcaatatcaacc
JW1579	atgcagctggcacgacaggtttcccgactgacgaaagtgattgcgcctacc
JW1580	ggattcctgaaacagaaagccg
JW1581	ctgctctgcggctttctgtttcaggaatcccagcgttgcgagtgcagtac
JW1582	tggcgggtgtcggggctggcttaactatgccactcgaggcgtttttcgttatg
JW1612	atgcagctggcacgacaggtttcccgactggatgctgaatcaatgatgtctgcc
JW1613	ggttttgtatttggggagcaatgg

JW1614	tcatcgccattgctccccaaatacaaaaccccagcgttgcgagtgcag tac
JW1619	tcatcgccattgctccccaaatacaaaaccatcctgaatgtaataagc gttcatgg
JW1620	tggcgggtgtcggggctggcttaactatgcacatcattgattacgact ggaaagc
JW2096	aggtcgcttatatggggatattctg
JW2097	ttgagattatggtgctgaccaaaag
JW2099	gtggtgattctatgattgatggtgg
JW3077	CAGTCATAGCCGAATAGCCT
JW3100	cggtagatttggatggtttaagg
JW3101	ggataatgctacctctggtgaagg
oAM201	gccccatacgatataagttgtaattccaaaccctatgctactccgt
oAM202	ccttaaacgcctggtgctaaacgcaaaaaggccatcc
oAM203	acggagtagcatagggtttggaattacaacttatatcgtatggggc
oAM204	ggatggccttttgcgtttagcaccaggcgtttaagg
hmdC_60_MfeI	TAGACATTGCCCTCGAGGTA(hmC)AATTGATCCGATTTCGACCTCAA ACCTAGACGAATTCCG
Bot_MfeI_60	CGGAATTCGTCTAGGTTTGAGGTCGAAATCGGATCAATTGTACCTCGA GGGCAATGTCTA
hmdC_60_MfeI_ Bot	CGGAATTCGTCTAGGTTTGAGGTCGAAATCGGAT(hmC)AATTGTACC TCGAGGGCAATGTCTA
dU_60	TAGACATTGCCCTCGAGGTAUCATGGATCCGATTTCGACCTCAAACCT AGACGAATTCCG
dT_60	TAGACATTGCCCTCGAGGTATCATGGATCCGATTTCGACCTCAAACCT AGACGAATTCCG
5mdC_60	TAGACATTGCCCTCGAGGTA(mC)CATGGATCCGATTTCGACCTCAAA CCTAGACGAATTCCG
2aminodA_60	TAGACATTGCCCTCGAGGTA(2-aminodA)CATGGATCCGATTTCGACCTCAAACCTAGACGAATTCCG
dU_60_2	TAGACATTGCCCTCGAGGTAUAATTGATCCGATTTCGACCTCAAACCT AGACGAATTCCG
Bot_dU_60_2	CGGAATTCGTCTAGGTTTGAGGTCGAAATCGGATCAATTATACCTCGA GGGCAATGTCTA
dU_60_2_Bot	CGGAATTCGTCTAGGTTTGAGGTCGAAATCGGAUCAATTATACCTCGA GGGCAATGTCTA
top_60_oligo	TAGACATTGCCCTCGAGGTACAATTGATCCGATTTCGACCTCAAACCT AGACGAATTCCG
top_40_oligo	CAATTGATCCGATTTCGACCTCAAACCTAGACGAATTCCG
bot_40_oligo	CGGAATTCGTCTAGGTTTGAGGTCGAAATCGGATCAATTG
top_20_oligo	TAGACATTGCCCTCGAGGTA
bot_20_oligo	TACCTCGAGGGCAATGTCTA
dU_18	TCGAGGTAUCATGGATCC
hmdC_18	TCGAGGTA(hmC)AATTGATCC

Table 9.2 Bacterial strains used in this study

Bacterial strain	Description	Source
E. coli EC100	TransforMax EC100 electrocompetent or chemically competent *E. coli*	Lucigen, now Biosearch Technologies (Cat# EC10010)
E. coli K-12 MG1655	common *E. coli* laboratory strain	Guyer et al., Cold Spring Harbor Symp Quant Biol, 1981
E. coli K-12 BW25113	parent strain of the Keio knockout collection at Coli Genetic Stock Center at Yale University; K-12 derivative	Datsenko & Wanner, PNAS, 2000
E. coli K-12 BW25113 Δ*xthA*	*xthA* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006
E. coli K-12 BW25113 Δ*nfo*	*nfo* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006
E. coli K-12 BW25113 Δ*nth*	*nth* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006
E. coli K-12 BW25113 Δ*yedK*	*yedK* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006
E. coli K-12 BW25113 Δ*recB*	*recB* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006
E. coli K-12 BW25113 Δ*recC*	*recC* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006
E. coli K-12 BW25113 Δ*recD*	*recD* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006
E. coli K-12 BW25113 Δ*recQ*	*recQ* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006

E. coli K-12 BW25113 Δ*recJ*	*recJ* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006
E. coli K-12 BW25113 Δ*recA*	*recA* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006
E. coli K-12 BW25113 Δ*polB*	*polB* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006
E. coli K-12 BW25113 Δ*dinB*	*dinB* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006
E. coli K-12 BW25113 Δ*umuC*	*umuC* gene knockout from Keio knockout collection at Coli Genetic Stock Center at Yale University; kanamycin selection marker	Baba et al., Mol Syst Biol, 2006
E. coli K-12 BW25113 Δ*xthA* Δ*nfo*	*xthA* and *nfo* gene double knockout made through λ Red recombineering of *nfo* Keio knockout collection strain above; tetracycline and kanamycin selection markers at *xthA* and *nfo* loci respectively	This study
E. coli ACT-01	*E. coli* K-12 MG1655 3xΔ:pACT-01 (chromosomally expresses type I-E CRISPR Cascade and Cas3 proteins from an arabinose-inducible promoter); obtained from Chris A. Voigt at MIT	Caliando and Voigt, Nat Commun, 2015
E. coli KD263	*E. coli* K-12 MG1655 strain that chromosomally expresses type I-E CRISPR Cascade, Cas1 and Cas2 proteins from an arabinose-inducible promoter and Cas3 from an IPTG-inducible promoter; obtained from Konstantin Severinov and Ekaterina Semenova at Rutgers University	Strotskaya et al., NAR 2017
E. coli λ_*E4-R*	KD263 with a λ-targeting type I-E CRISPR spacer acquired from λ DNA cloned onto a plasmid; obtained from Konstantin Severinov and Ekaterina Semenova at Rutgers University	Strotskaya et al., NAR 2017
E. coli λ_*L1-R*	KD263 with a λ-targeting type I-E CRISPR spacer acquired from λ DNA cloned onto a plasmid; obtained from Konstantin Severinov and Ekaterina Semenova at Rutgers University	Strotskaya et al., NAR 2017
E. coli λ_*L4-R*	KD263 with a λ-targeting type I-E CRISPR spacer acquired from λ DNA cloned onto a plasmid; obtained from Konstantin Severinov and Ekaterina Semenova at Rutgers University	Strotskaya et al., NAR 2017
E. coli λ_*L6-R*	KD263 with a λ-targeting type I-E CRISPR spacer acquired from λ DNA cloned onto a plasmid; obtained	Strotskaya et al., NAR 2017

	from Konstantin Severinov and Ekaterina Semenova at Rutgers University	
E. coli Δ*polB* λ_*L6-R*	*E. coli* λ_*L6-R* with a *polB* gene deletion (substituted with a KanR selectable marker); same deletion as *polB* gene knockout from Keio knockout collection	This study
E. coli Δ*dinB* λ_*L6-R*	*E. coli* λ_*L6-R* with a *dinB* gene deletion (substituted with a KanR selectable marker); same deletion as *dinB* gene knockout from Keio knockout collection	This study
E. coli Δ*umuC* λ_*L6-R*	*E. coli* λ_*L6-R* with a *umuC* gene deletion (substituted with a KanR selectable marker); same deletion as *umuC* gene knockout from Keio knockout collection	This study
E. coli K-12 MG1655 Δ9	*E. coli* strain with all 9 cryptic prophages deleted; obtained from Thomas K. Wood at Penn State University	Wang et al., Nat Commun, 2010
E. coli K-12 MG1655::λ	A lysogen of bacteriophage λ	JW Roberts
E. coli K-12 MG1655::λ ΔP	A lysogen of bacteriophage λΔP	This study

Table 9.3 Plasmids and cosmids used in this study

Plasmid/cosmid	Description	Source	Construction Notes
pCas9::*spcNT*	pACYC184 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and BsaI spacer for oligo cloning	pCas9 in Jiang et al., Nat Biotechnol, 2013	
pCas9::*spc9*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc9* targeting phage λ*vir*	This study	BsaI cloning into pCas9::*spcNT* with AA318/AA319
pCas9::*spc40*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc40* targeting phage λ*vir*	This study	BsaI cloning into pCas9::*spcNT* with AA132/AA133
pCas9::*spc45*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc45* targeting phage λ*vir*	This study	BsaI cloning into pCas9::*spcNT* with AA153/AA154
pCas9::*spc45c*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc45c* targeting phage λ*vir*	This study	BsaI cloning into pCas9::*spcNT* with AA383/AA384
pCas9::*spc14*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc14* targeting phage λ*vir*	This study	BsaI cloning into pCas9::*spcNT* with AA320/AA321
pCas9::*spc15*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc15* targeting phage λ*vir*	This study	BsaI cloning into pCas9::*spcNT* with AA349/AA350
pCas9::*spc26D*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc26D* targeting phage λ*vir*	This study	BsaI cloning into pCas9::*spcNT* with AA107/AA108
pdCas9::*spcNT*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR dCas9 (D10A, H840A), tracrRNA and BsaI spacer for oligo cloning	pdCas9 in Bikard et al., NAR, 2013	
pdCas9::*spc40*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR dCas9 (D10A, H840A), tracrRNA and *spc40* targeting phage λ*vir*	This study	BsaI cloning into pdCas9::*spcNT* with AA132/AA133
pdCas9::*spc26D*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR dCas9 (D10A,	This study	BsaI cloning into pdCas9::*spcNT* with AA107/AA108

	H840A), tracrRNA and *spc26D* targeting phage λ*vir*		
pBAD18	plasmid with P_{ara} promoter and *araC* gene; P_{ara} is an arabinose inducible promoter; used to make pAM38 plasmid	Guzman et al., J Bacteriol, 1995	
pPL2e	plasmid with p15A origin of replication and *cat* gene; used to make pAM38 plasmid	Lauer et al., J Bacteriol, 2002	
pAM38	P_{ara} *cat* p15A ori plasmid; Para is an arabinose inducible promoter	This study	Gibson assembly from pBAD18 (oAM201/oAM202) and pPL2E (oAM203/oAM204)
pAM38(*gam*)	pAM38 plasmid with λ Gam protein being expressed under Para promoter	This study	Gibson assembly from pAM38 (JW1484/JW1485) and λ*vir* genomic DNA (JW1567/JW1568)
pAM38(*exo*)	pAM38 plasmid with λ Exo protein being expressed under Para promoter	This study	Gibson assembly from pAM38(*gam*) (JW1484/AA149) and λ*vir* genomic DNA (AA147/AA148)
pAM38(*bet*)	pAM38 plasmid with λ Beta protein being expressed under Para promoter	This study	Gibson assembly from pAM38(*gam*) (JW1484/AA149) and λ*vir* genomic DNA (AA145/AA146)
pDS-SPcas	cloDF13 ori spectinomycin-resistant plasmid expressing *S. pyogenes* Cas9, tracrRNA and a single repeat; used to make pDS-SPcas9::*spcNT*	Addgene plasmid 48645 (Esvelt et al., Nat Methods, 2013)	
pJM62	pC194-based medium-copy plasmid expressing *S. pyogenes* type II-A CRISPR-Cas system with two repeats flanking BsaI sites for spacer cloning with oligos; used to make pDS-SPcas9::*spcNT*	Modell et al., Nature, 2017	
pDS-SPcas9::*spcNT*	pDS plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and BsaI spacer for oligo cloning; spectinomycin resistance	This study	Gibson assembly from pDS-SPcas (JW488/JW489) and pJM62 (JW1188/JW1189)

pDS-SPcas9::*spc9*	pDS plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc9* targeting phage λ*vir*; spectinomycin resistance	This study	BsaI cloning into pDS-SPcas9::*spcNT* with AA318/AA319
pDS-SPcas9::*spcNT*(CmR)	pDS plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and BsaI spacer for oligo cloning; chloramphenicol resistance	This study	Gibson assembly from pDS-SPcas9::*spcNT* (AA68/AA69) and pCas9::*spcNT* (AA66/AA67)
pDS-SPcas9::*spc9*(CmR)	pDS plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc9* targeting phage λ*vir*; chloramphenicol resistance	This study	BsaI cloning into pDS-SPcas9::*spcNT*(CmR) with AA318/AA319
pEmpty (pCL1920)	a pSC101-derived low-copy number plasmid with spectinomycin/strepomycin resistance	Lerner and Inouye, NAR, 1990	
pRecB	pCL1920 plasmid expressing *E. coli* K-12 MG1655 RecB protein	This study	Gibson assembly from pCL1920 (JW1209/JW1059) and *E. coli* K-12 MG1655 genomic DNA (JW1522/JW1523)
pRecB(D1080A)	pCL1920 plasmid expressing *E. coli* K-12 MG1655 RecB protein with D1080A mutation in the RecB nuclease domain	This study	Gibson assembly from pCL1920 (JW1209/JW1059) and *E. coli* K-12 MG1655 genomic DNA (JW1522/JW1524 and JW1525/JW1523)
pDinB	pCL1920 plasmid expressing *E. coli* K-12 MG1655 DinB protein	This study	Gibson assembly from pCL1920 (AA696/AA697) and *E. coli* K-12 MG1655 genomic DNA (AA694/AA695)
pKM208(*red*)	pKM208 plasmid from Kenan C. Murphy (UMass Amherst) with λ *red* genes under Ptac IPTG-inducible promoter	Murphy and Campellone, BMB Mol Biol, 2003	
pKM208(empty)	empty pKM208 plasmid from Kenan C. Murphy (UMass Amherst) with λ *red* genes cloned out	This study	Gibson assembly from pKM208(*red*) (AA333/AA331 and AA327/AA332)

pPD207.842	plasmid from Andrew Z. Fire with *E. coli* K-12 MG1655 type I-E CRISPR BsaI spacer for oligo cloning; spacer expression under Para promoter	Fu et al., Genetics, 2017	
pACYC184-TypeIE*spcNT*	pACYC184 plasmid with *E. coli* K-12 MG1655 type I-E CRISPR BsaI spacer for oligo cloning; spacer expression under Para promoter	This study	Gibson assembly from pCas9::*spcNT* (AA396/AA397) and pPD207.842 (AA398/AA399)
pACYC184-TypeIE*spc9R*	pACYC184 plasmid with *E. coli* K-12 MG1655 type I-E CRISPR spacer *spc9R* targeting phage λ*vir*; spacer expression under Para promoter	This study	BsaI cloning into pACYC184-TypeIE*spcNT* with AA77/AA78
pUT18C	high-copy pUC19-derived plasmid	Karimova et al., PNAS, 1998	
pUT18C-d*gam*	pUT18C plasmid with cloned segment of lambda *red* region with *gam* deletion for phage recombination	This study	Gibson assembly from pUT18C (JW1409/JW1410) and λ*vir* Δ*gam*2 genomic DNA (JW1554/JW1555)
pUT18C-d*exo*	pUT18C plasmid with cloned segment of lambda *red* region with *exo* deletion for phage recombination	This study	Gibson assembly from pUT18C (JW1409/JW1410) and λ*vir* Δ*exo*4 genomic DNA (JW1556/JW1557)
pUT18C-d*bet*	pUT18C plasmid with cloned segment of lambda *red* region with *bet* deletion for phage recombination	This study	Gibson assembly from pUT18C (JW1409/JW1410) and λ*vir* genomic DNA (JW1579/JW1580 and JW1581/JW1582)
pUT18C-d*exod**bet*	pUT18C plasmid with cloned segment of lambda *red* region with *bet* and *exo* deletions for phage recombination	This study	Gibson assembly from pUT18C (JW1409/JW1410) and λ*vir* genomic DNA (JW1612/JW1613 and JW1614/JW1582)
pUT18C-d*red*	pUT18C plasmid with cloned segment of lambda *red* region	This study	Gibson assembly from pUT18C

	with *gam*, *bet* and *exo* full *red* operon deletion for phage recombination		(JW1409/JW1410) and λ*vir* genomic DNA (JW1612/JW1613 and JW1619/JW1620)
pUT18C-*chi*D	pUT18C plasmid with cloned segment of lambda cryptic *chi* site *chi*D converted to a functional *E. coli chi* site for phage recombination	This study	Gibson assembly from pUT18C (JW1409/JW1410) and λ*vir* genomic DNA (JW1548/JW1549 and JW1550/JW1551)
pUT18C-3*chi*F+R	pUT18C plasmid with cloned segment of lambda *spc14* target region with three tandem *chi* sites on the top strand and three tandem chi sites on the bottom strand surrounding protospacer 14 for phage recombination	This study	Gibson assembly from pUT18C (JW1409/JW1410) and λ*vir* genomic DNA (JW1413/JW1561, JW1562/JW1377, JW1417/JW1491, and JW1406/JW1418)
pCas9::*spc12*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc12* targeting phage λ*vir exo* for selection of *exo* mutant phages during phage construction	This study	BsaI cloning into pCas9::*spcNT* with JW1191/JW1192
pCas9::JW1370-1	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *a spacer* targeting phage λ*vir gam* for selection of *gam* mutant phages during phage construction	This study	BsaI cloning into pCas9::*spcNT* with JW1370/JW1371
pCas9::JW1552-3	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *a spacer* targeting phage λ*vir bet* for selection of *bet* mutant phages during phage construction	This study	BsaI cloning into pCas9::*spcNT* with JW1552/JW1553
pCas9::JW1546-7	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *a spacer* targeting phage λ*vir* cryptic *chi*D for selection of	This study	BsaI cloning into pCas9::*spcNT* with JW1546/JW1547

	*chi*D (*chi*1) mutant phages during phage construction		
pCas9::JW1558-9	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *a spacer* targeting phage λ*vir* for selection of +*3chi*F (*chi*2-4) mutant phages during phage construction	This study	BsaI cloning into pCas9::*spcNT* with JW1558/JW1559
pCas9::JW1401-2	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *a spacer* targeting phage λ*vir* for selection of +*3chi*R (*chi*5-7) mutant phages during phage construction	This study	BsaI cloning into pCas9::*spcNT* with JW1401/JW1402
pKOBEG-A	a Ts replication pSC101 derivative expressing *gam*, *bet*, and *exo* from the arabinose-inducible pBAD promoter	Chaveroche et al., NAR, 2000	
pKD4	an R6K ori plasmid with a kanamycin-resistance gene	Datsenko and Wanner, PNAS, 2000	
pCL1920-P	pCL1920 plasmid expressing lambda gene *P* from the *lac* promoter	This study	Gibson assembly from pCL1920 (JW1209/JW1059) and λ*vir* genomic DNA (JW1480/JW1481)
pWEB-TNC	commericallly available cloning cosmid with pBR322 origin, ampicllin resistance and chloramphenicol resistance	EPICENTRE Biotechnologies	
anti-T4 resistant cosmid	metagenomic cosmid isolated from AZ52 megapool 4 library after three rounds of T4 phage selection; metagenomic DNA is cloned in pWEB-TNC cosmid between T7 and M13-F40 primer sites	This study	
pFragmentC	Fragment C from anti-T4 resistant cosmid cloned into pWEB-TNC	This study	NEBuilder® HiFi DNA assembly from pWEB-TNC (AA559/AA560) and anti-T4 resistant cosmid (AA627/AA629)

pFragmentD	Fragment D from anti-T4 resistant cosmid cloned into pWEB-TNC	This study	NEBuilder® HiFi DNA assembly from pWEB-TNC (AA559/AA560) and anti-T4 resistant cosmid (AA615/AA625)
pFragmentD1	Fragment D1 from pFragment D cloned into pWEB-TNC	This study	NEBuilder® HiFi DNA assembly from pWEB-TNC (AA559/AA560) and pFragmentD (AA615/AA640)
pFragmentD2	Fragment D2 from pFragment D cloned into pWEB-TNC	This study	NEBuilder® HiFi DNA assembly from pWEB-TNC (AA559/AA560) and pFragmentD (AA643/AA644)
pFragmentD3	Fragment D3 from pFragment D cloned into pWEB-TNC	This study	NEBuilder® HiFi DNA assembly from pWEB-TNC (AA559/AA560) and pFragmentD (AA630/AA646)
pFragmentD3-1	gene *a* from three-gene operon in Fragment D3 cloned into pWEB-TNC	This study	NEBuilder® HiFi DNA assembly from pFragmentD3 (AA559/AA560 and AA646/AA676)
pFragmentD3-2	gene *b* from three-gene operon in Fragment D3 cloned into pWEB-TNC	This study	NEBuilder® HiFi DNA assembly from pFragmentD3 (AA559/AA674 and AA672/AA673)
pFragmentD3-3	genes *b*+*c* from three-gene operon in Fragment D3 cloned into pWEB-TNC	This study	NEBuilder® HiFi DNA assembly from pFragmentD3 (AA559/AA674 and AA625/AA673)
pFragmentD3-4 (pBrig1)	gene *c* (*brig1*) from three-gene operon in Fragment D3 cloned into pWEB-TNC	This study	NEBuilder® HiFi DNA assembly from pFragmentD3 (AA559/AA674 and AA630/AA675)
pFragmentD3-5	genes *a*+*b* from three-gene operon in Fragment D3 cloned into pWEB-TNC	This study	NEBuilder® HiFi DNA assembly from pFragmentD3 (AA559/AA560 and AA646/AA672)

pFragmentD3-6	genes *a+c* from three-gene operon in Fragment D3 cloned into pWEB-TNC	This study	NEBuilder® HiFi DNA assembly from pFragmentD3 (AA559/AA677and AA625/AA678)
pAM38(*brig1*)	pAM38 plasmid with Brig1 being expressed under arabinose-inducible Para promoter using λ Gam RBS; Para cat p15A ori plasmid	This study	Gibson assembly from pAM38(*gam*) (AA150/AA693) and pFragmentD3-4 (AA816/AA817)
pAM38(*red*)	pAM38 plasmid with λ Gam, Beta and Exo protein being expressed under arabinose-inducible Para promoter using λ Gam RBS; Para cat p15A ori plasmid	This study	Gibson assembly from pAM38(*gam*) (AA149/AA150) and and λ*vir* genomic DNA (AA148/AA207)
p(*a-gt*)	pCL1920 plasmid expressing bacteriophage T4 *a-gt* gene which encodes alpha-glucosyltransferase (a-GT)	This study	Gibson assembly from pRecB (AA246/AA903) and bacteriophage T4 genomic DNA (AA901/AA902)
p(*b-gt*)	pCL1920 plasmid expressing bacteriophage T4 *a-gt* gene which encodes beta-glucosyltransferase (b-GT)	This study	Gibson assembly from pRecB (AA246/AA903) and bacteriophage T4 genomic DNA (AA904/AA905)
pAM38(*brig1*)-SmR	pAM38 plasmid with Brig1 being expressed under arabinose-inducible Para promoter using λ Gam RBS; Para p15A ori plasmid with spectinomycin resistance	This study	Gibson assembly from pAM38(*brig1*) (AA990/AA991) and p(*a-gt*) (AA988/AA989)
p(*gp42*)	pAM38 plasmid with bacteriophage T4 Gp42 hydroxymethylase being expressed under arabinose-inducible Para promoter using λ Gam RBS; Para cat p15A ori plasmid with spectinomycin resistance	This study	Gibson assembly from pAM38(*brig1*)-SmR (AA150/AA693) and bacteriophage T4 genomic DNA (AA1104/AA1105)
pET21a	protein expression vector with ampicillin/carbenicillin resistance	Addgene 69740-3	
pET21a-Brig1-6xHis	Brig1 with a C-terminal 6xHis tag cloned into pET21a plasmid vector for protein purification	This study	Gibson assembly from pET21a (AA1320/AA1321) and pBrig1 (AA1331/AA1332)

pET21b-6xHis-a-GT	bacteriophage T4 *a-gt* gene which encodes alpha-glucosyltransferase (a-GT) with an N-terminal 6xHis tag cloned into pET21b plasmid vector for protein purification	Gift from Joshua S. Chappie at Cornell University	
pBrig1(Y121A)	pBrig1 with Y121A mutation in Brig1	This study	Gibson assembly from pFragmentD3-4 (AA559/AA956 and AA614/AA957)
pBrig1(N145A)	pBrig1 with N145A mutation in Brig1	This study	Gibson assembly from pFragmentD3-4 (AA559/AA960 and AA614/AA961)
pBrig1(E147A)	pBrig1 with E147A mutation in Brig1	This study	Gibson assembly from pFragmentD3-4 (AA559/AA954 and AA614/AA955)
pBrig1(E147Q)	pBrig1 with E147Q mutation in Brig1	This study	Gibson assembly from pFragmentD3-4 (AA559/AA958 and AA614/AA959)
pET21a-Brig1(Y121A)-6xHis	Brig1(Y121A) mutant with a C-terminal 6xHis tag cloned into pET21a plasmid vector for protein purification	This study	Gibson assembly from pET21a (AA1320/AA1321) and pBrig1(Y121A) (AA1331/AA1332)
pET21a-Brig1(Y121A,E147A)-6xHis	Brig1(Y121A, E147A) double mutant with a C-terminal 6xHis tag cloned into pET21a plasmid vector for protein purification	This study	Gibson assembly from pET21a-Brig1(Y121A)-6xHis (AA955/AA1321) and pET21a-Brig1(Y121A)-6xHis (AA954/AA1331)
pAM38(*xthA*)	pAM38 plasmid with *E. coli* Exonuclease III (XthA) being expressed under arabinose-inducible Para promoter using λ Gam RBS; Para cat p15A ori plasmid with spectinomycin resistance	This study	Gibson assembly from pAM38(*brig1*)-SmR (AA150/AA693) and *E. coli* MG1655 genomic DNA (AA994/AA995)
pAM38(*nfo*)	pAM38 plasmid with *E. coli* Endonuclease IV (Nfo) being expressed under arabinose-inducible Para promoter using λ Gam RBS; Para cat p15A ori plasmid with spectinomycin resistance	This study	Gibson assembly from pAM38(*brig1*)-SmR (AA150/AA693) and *E. coli* MG1655 genomic DNA (AA992/AA993)
pAM39	same as pAM38, but with increased spacing between	Gift from Alex J.	

	ATG start site of gene and with a canonical *E. coli* consensus RBS sequence; provides higher expression under arabinose induction	Meeske at University of Washington, Seattle	
pAM39(WP129427366.1)	pAM39 vector overexpressing Brig1 homolog from *Nocardioides zhouii* (NCBI WP129427366.1) under arabinose-inducible Para promoter	This study	Gibson assembly from pAM39 (AA1020/AA1021) and synthesized IDT gBlock of WP129427366.1 gene sequence with 30 bp overhangs for direct cloning
pAM39(WP165228961.1)	pAM39 vector overexpressing Brig1 homolog from *Nocardioides anomalu*s (NCBI WP165228961.1) under arabinose-inducible Para promoter	This study	Gibson assembly from pAM39 (AA1020/AA1021) and synthesized IDT gBlock of WP165228961.1 gene sequence with 30 bp overhangs for direct cloning
pUT18C-d*a-gt*	pUT18C plasmid with cloned segment of T4 genomic DNA comprising the region surrounding the *a-gt* gene with an inframe *a-gt* deletion for phage recombination	This study	Gibson assembly from pUT18C (AA219/AA220) and bacteriophage T4 genomic DNA (AA922/AA923 and AA924/AA925)
pUT18C-d*b-gt*	pUT18C plasmid with cloned segment of T4 genomic DNA comprising the region surrounding the *b-gt* gene with an inframe *b-gt* deletion for phage recombination	This study	Gibson assembly from pUT18C (AA219/AA220) and bacteriophage T4 genomic DNA (AA932/AA933 and AA934/AA935)
pUT18C-d*alc*	pUT18C plasmid with cloned segment of T4 genomic DNA comprising the region surrounding the *alc* gene with an inframe *alc* deletion for phage recombination	This study	Gibson assembly from pUT18C (AA219/AA220) and bacteriophage T4 genomic DNA (AA962/AA963 and AA964/AA965)
pUT18C-d*denB*	pUT18C plasmid with cloned segment of T4 genomic DNA comprising the region surrounding the *denB* gene with an inframe *denB* deletion for phage recombination	This study	Gibson assembly from pUT18C (AA219/AA220) and bacteriophage T4 genomic DNA

			(AA893/AA894 and AA895/AA896)
pUT18C-d*gp56*	pUT18C plasmid with cloned segment of T4 genomic DNA comprising the region surrounding the *gp56* gene with an inframe *gp56* deletion for phage recombination	This study	Gibson assembly from pUT18C (AA219/AA220) and bacteriophage T4 genomic DNA (AA897/AA898 and AA899/AA900)
pUT18C-d*gp42*	pUT18C plasmid with cloned segment of T4 genomic DNA comprising the region surrounding the *gp42* gene with an inframe *gp42* deletion for phage recombination	This study	Gibson assembly from pUT18C (AA219/AA220) and bacteriophage T4 genomic DNA (AA906/AA907 and AA908/AA909)
pUT18C-d*ba-gt*	pUT18C plasmid with cloned segment of T6 genomic DNA comprising the region surrounding the *ba-gt* gene with an inframe *ba-gt* deletion for phage recombination	This study	Gibson assembly from pUT18C (AA219/AA220) and bacteriophage T6 genomic DNA (AA1171/AA1172 and AA1173/AA1174)
pCas9::*spc-bgt-c*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc-bgt-c* targeting phage T4 for selection of Δ*b-gt* mutant phages during phage construction	This study	BsaI cloning into pCas9::*spcNT* with AA940/AA941
pCas9::*spc-gp56-7*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc-gp56-7* targeting phage T4 for selection of Δ*gp56* mutant phages during phage construction	This study	BsaI cloning into pCas9::*spcNT* with AA1068/AA1069
pCas9::*spc-gp42-3*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc-gp42-3* targeting phage T4 for selection of Δ*gp42* mutant phages during phage construction	This study	BsaI cloning into pCas9::*spcNT* with AA1088/AA1089
pCas9::*spc-gp42-4*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc-gp42-4* targeting phage T4 for selection of Δ*gp42* mutant	This study	BsaI cloning into pCas9::*spcNT* with AA1090/AA1091

	phages during phage construction		
pCas9::*spc-ba-gt-10*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc-ba-gt-10* targeting phage T6 for selection of Δ*ba-gt* mutant phages during phage construction	This study	BsaI cloning into pCas9::*spcNT* with AA1196/AA1197
pCas9::*spc-ba-gt-11*	pCas9 plasmid with *S. pyogenes* SF370 type II-A CRISPR Cas9, tracrRNA and *spc-ba-gt-11* targeting phage T6 for selection of Δ*ba-gt* mutant phages during phage construction	This study	BsaI cloning into pCas9::*spcNT* with AA1198/AA1199
pACYC-TypeIE*spc-alc-1*	pACYC184 plasmid with *E. coli* K-12 MG1655 type I-E CRISPR spacer *spc-alc-1* targeting phage T4 for selection of Δ*alc* mutant phages during phage construction; spacer expression under Para promoter	This study	BsaI cloning into pACYC184-TypeIE*spcNT* with AA974/AA975
pACYC-TypeIE*spc-alc-2*	pACYC184 plasmid with *E. coli* K-12 MG1655 type I-E CRISPR spacer *spc-alc-2* targeting phage T4 for selection of Δ*alc* mutant phages during phage construction; spacer expression under Para promoter	This study	BsaI cloning into pACYC184-TypeIE*spcNT* with AA976/AA977
pACYC-TypeIE*spc-alc-3*	pACYC184 plasmid with *E. coli* K-12 MG1655 type I-E CRISPR spacer *spc-alc-3* targeting phage T4 for selection of Δ*alc* mutant phages during phage construction; spacer expression under Para promoter	This study	BsaI cloning into pACYC184-TypeIE*spcNT* with AA978/AA979
pACYC-TypeIE*spc-denB-1*	pACYC184 plasmid with *E. coli* K-12 MG1655 type I-E CRISPR spacer *spc-denB-1* targeting phage T4 for selection of Δ*denB* mutant phages during phage construction; spacer	This study	BsaI cloning into pACYC184-TypeIE*spcNT* with AA1002/AA1003

	expression under Para promoter		
pACYC-TypeIE*spc-denB-2*	pACYC184 plasmid with *E. coli* K-12 MG1655 type I-E CRISPR spacer *spc-denB-2* targeting phage T4 for selection of Δ*denB* mutant phages during phage construction; spacer expression under Para promoter	This study	BsaI cloning into pACYC184-TypeIE*spcNT* with AA1004/AA1005

Table 9.4 CRISPR spacers used in λ Red experiments

CRISPR type/species	Spacer	Sequence (5'-3')
II-A from *S. pyogenes* SF370	*spcNT*	TGAGACCAGTCTCGGAAGCTCAAAGGTCTC
II-A from *S. pyogenes* SF370	*spc9*	AGATACCAGCTTCACGCTGGCGTGGATGCC
II-A from *S. pyogenes* SF370	*spc40*	AGGGATGCACCATTCTGAGATGTTTTTATT
II-A from *S. pyogenes* SF370	*spc45*	ACTCGTCAGAATGAATATTATCAAGCAGCA
II-A from *S. pyogenes* SF370	*spc45c*	ACTCGTCAGGATGAATATTTTCAAGCAGCA
II-A from *S. pyogenes* SF370	*spc14*	CAGGGGTGTTACCACTACCGCAGGAAAAGG
II-A from *S. pyogenes* SF370	*spc15*	CACCATCAGTTCAAGACGACGCAGCACCTC
II-A from *S. pyogenes* SF370	*spc26D*	TACGGCGCAATTCCGCATCAGTAAGCGCAT
I-E from *E-coli* K-12 MG1655	*spc9R*	CACGCCAGCAGCGCCTGCTGCCCCTGCTCTCC
I-E from *E-coli* K-12 MG1655	*spcE4-R*	ATAACGCTTGTGAAAATGCTGAATTTCGCGTC
I-E from *E-coli* K-12 MG1655	*spcL1-R*	GGTGGGAATGGTGGGCGTTTTCATACATAAAA
I-E from *E-coli* K-12 MG1655	*spcL4-R*	CTGCTCATACGAGACACCCAGCCCGGCAGCGA
I-E from *E-coli* K-12 MG1655	*spcL6-R*	ACCGGCTGCACGGCGCTCCATCGTTTCACGGA

Table 9.5 Bacteriophages used in this study

Phage	Description	Plasmids used for construction	Source
λ*vir* parent stock	a non-lysogenic variant of λ (*Siphoviridae*) used to make all subsequent phage stocks		Bruce Levin
λ*vir* Δ*gam*2	λ*vir* escaper of pCas9:JW1370-1 with a 144 bp deletion in *gam* (deleted bases: 26,883-27,026)	pCas9:JW1370-1	This study
λ*vir* Δ*exo*4	λ*vir* escaper of pCas9:*spc12* with a 74 bp deletion in *exo* (deleted bases: 25,827-25,900)	pCas9:*spc12*	This study
λ*vir*	a non-lysogenic variant of λ used in experiments (parent stock above passaged through an empty vector strain and a Cas9 non-targeting strain)	pUT18C, pCas9::*spcNT*	This study
λ*vir* Δ*gam*	a non-lysogenic variant of λ with a 144 bp deletion in *gam* (deleted bases: 26,883-27,026)	pUT18C-D*gam,* pCas9:JW1370-1	This study
λ*vir* Δ*exo*	a non-lysogenic variant of λ with a 74 bp deletion in *exo* (deleted bases: 25,827-25,900)	pUT18C-D*exo,* pCas9:*spc12*	This study
λ*vir* Δ*bet*	a non-lysogenic variant of λ with a 717 bp deletion in *bet* (deleted bases: 26,075-26,791)	pUT18C-D*bet,* pCas9:JW1552-3	This study
λ*vir* Δ*exo* Δ*bet*	a non-lysogenic variant of λ with a 1408 bp deletion spanning *exo* and *bet* (deleted bases: 25,383-26,790)	pUT18C-D*exo*D*bet,* pCas9:JW1552-3	This study
λ*vir* Δ*red*	a non-lysogenic variant of λ with a 1711 bp deletion spanning *gam*, *exo* and *bet* (deleted bases: 25,383-27,093)	pUT18C-D*red*, pCas9:JW1552-3	This study
λ*vir chi*^{1} parent stock	a non-lysogenic variant of λ with a G(39,031)T mutation, creating a *chi* site, *chi*^{1}; used to make subsequent phage stocks	pUT18C-*chi*D, pCas9:JW1546-7	This study
λ*vir chi*^{1}	a non-lysogenic variant of λ with a G(39,031)T mutation, creating a *chi* site, *chi*^{1} (parent stock passaged through an empty vector strain and a Cas9 non-targeting strain)	pUT18C-*chi*D, pCas9:JW1546-7, pUT18C, pCas9::*spcNT*	This study
λ*vir chi*^{1} Δ*gam*	a variant of λ*vir chi*^{1} with a 144 bp deletion in *gam* (deleted bases: 26,883-27,026)	pUT18C-*chi*D, pCas9:JW1546-7, pUT18C-D*gam,* pCas9:JW1370-1	This study
λ*vir chi*^{1} Δ*exo*	a variant of λ*vir chi*^{1} with a 74 bp deletion in *exo* (deleted bases: 25,827-25,900)	pUT18C-*chi*D, pCas9:JW1546-7, pUT18C-D*exo,* pCas9:*spc12*	This study

λ*vir chi*[1] Δ*bet*	a variant of λ*vir chi*[1] with a 717 bp deletion in *bet* (deleted bases: 26,075-26,791)	pUT18C-*chi*D, pCas9:JW1546-7, pUT18C-D*bet,* pCas9:JW1552-3	This study
λ*vir chi*[1] Δ*exo*Δ*bet*	a variant of λ*vir chi*[1] with a 1408 bp deletion spanning *exo* and *bet* (deleted bases: 25,383-26,790)	pUT18C-*chi*D, pCas9:JW1546-7, pUT18C-D*exo*D*bet,* pCas9:JW1552-3	This study
λ*vir chi*[1] Δ*red*	a variant of λ*vir chi*[1] with a 1711 bp deletion spanning *gam*, *exo* and *bet* (deleted bases: 25,383-27,093)	pUT18C-*chi*D, pCas9:JW1546-7, pUT18C-D*red*, pCas9:JW1552-3	This study
λ*vir chi*[2-7] parent stock	a non-lysogenic variant of λ with 3 tandem *chi* sites (GCTGGTGGTCGATGCTGAGCTGGTGGACACGCGCTGGCTGGTGG), each separated by 10 bp, inserted in the forward direction replacing bases 35,777-35,809 (3*chi*F) and in the reverse direction replacing bases 36,328-36,360 (3*chi*R)	pUT18C-3*chi*F+R, pCas9:JW1558-9, pCas9:JW1401-2	This study
λ*vir chi*[2-7]	parent stock above passaged through an empty vector strain and a Cas9 non-targeting strain	pUT18C-3*chi*F+R, pCas9:JW1558-9, pCas9:JW1401-2, pUT18C, pCas9::*spcNT*	This study
λ*vir chi*[2-7] Δ*red*	a variant of λ*vir chi*[2-7] with a 1711 bp deletion spanning *gam*, *exo* and *bet* (deleted bases: 25,383-27,093)	pUT18C-3*chi*F+R, pCas9:JW1558-9, pCas9:JW1401-2, pUT18C-D*red*, pCas9:JW1552-3	This study
λΔP	a λ mutant containing a deletion of gene *P* (retaining the P start codon and last 6 amino acids)	pKD4, pKOBEG-A, pCL1920-P	This study
P1	*E. coli* bacteriophage P1		ATCC 25404-B1
T2	T-even lytic coliphage T2 (*Myoviridae*)		ATCC 11303-B2
T3	lytic coliphage T3 (*Autographiviridae*)		ATCC 11303-B3
T4	T-even lytic coliphage T4 (*Myoviridae*); used to construct T4 mutant phages below		Bruce Levin
T5	lytic coliphage T5 (*Demerecviridae*)		ATCC 11303-B5

T6	T-even lytic coliphage T6 (*Myoviridae*); used to construct T6 mutant phages below		ATCC 11303-B6
T7	lytic coliphage T7 (*Autographiviridae*)		Bruce Levin
T4 escaper1	mutant T4 phage isolated from a lawn of pBrig1-carrying *E. coli* EC100 cells with a 1 bp deletion of nt 713 (frameshift) in *a-gt* gene	pBrig1	This study
T4 Δ*a-gt*	T4 mutant phage with an inframe deletion of *a-gt*; phage does not encode alpha-glucosyltransferase (a-GT), only has beta-gluocosyl-hmC nucleobases	pUT18C-d*a-gt*, pBrig1	This study
T4 Δ*b-gt*	T4 mutant phage with an inframe deletion of *b-gt*; phage does not encode beta-glucosyltransferase (b-GT), only has alpha-gluocosyl-hmC nucleobases	pUT18C-d*b-gt*, pCas9::*spc-bgt-c*	This study
T4 Δ*alc* Δ*denB* Δ*gp56*	T4 mutant phage with inframe deletions of *alc*, *denB* and *gp56*, with deletions generated in that order; used to construct phage T4(C)	pUT18C-d*alc*, pACYC-TypeIE*spc-alc-1*, pACYC-TypeIE*spc-alc-2*, pACYC-TypeIE*spc-alc-3*, pUT18C-d*denB*, pACYC-TypeIE*spc-denB-1*, pACYC-TypeIE*spc-denB-2*, pUT18C-d*gp56*, pCas9::*spc-gp56-7*	This study
T4(C)	T4 Δ*alc* Δ*denB* Δ*gp56* Δ*gp42;* T4 mutant phage from above with an additional inframe deletion in gene *gp42*; phage lacks hmC nucleobases, carries cytosines instead	pUT18C-d*gp42*, pCas9::*spc-gp42-3*, pCas9::*spc-gp42-4*	This study
T4(+b-GT)	T4 phage passaged through *E. coli* EC100 cells carrying p(*b-gt*), overexpressing T4 b-GT under 1 mM IPTG induction to increase fraction of beta-gluocosyl-hmC nucleobases in T4 genome	p(*b-gt*)	This study
T6 escaper1	mutant T6 phage isolated from a lawn of pBrig1-carrying *E. coli* EC100 cells with nt 716 C>T mutation (Ala239->Val239) in *a-gt* gene	pBrig1	This study
T6 escaper2	mutant T6 phage isolated from a lawn of pBrig1-carrying *E. coli* EC100 cells with a	pBrig1	This study

	1 bp insertion (C) after nt 758 (frameshift) in *a-gt* gene		
T6 Δ*ba-gt*	T6 mutant phage with an inframe deletion of *ba-gt*; phage does not encode beta-alpha glucosyltransferase (ba-GT), only has alpa-gluocosyl-hmC nucleobases and no gentiobiosyl-hmC nucleobases	pUT18C-d*ba-gt*, pCas9::*spc-ba-gt-10*, pCas9::*spc-ba-gt-11*	This study
Bas35-45	T-even phages of *Tequatrovirus* subfamily from BASEL collection with alpha-glucosyl-hmC nucleobases		Maffei et al., PLoS Biol, 2021
Bas46-47	T-even phages of *Mosigvirus* subfamily from BASEL collection with arabinosyl-hmC nucleobases		Maffei et al., PLoS Biol, 2021

CHAPTER 10. REFERENCES

1. Ablasser, A., Goldeck, M., Cavlar, T., Deimling, T., Witte, G., Rohl, I., Hopfner, K.P., Ludwig, J., and Hornung, V. (2013). cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. Nature *498*, 380-384.
2. Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L.*, et al.* (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science *353*, aaf5573.
3. Ackermann, H.W. (2007). 5500 Phages examined in the electron microscope. Arch Virol *152*, 227-243.
4. Ahn, W.C., Aroli, S., Kim, J.H., Moon, J.H., Lee, G.S., Lee, M.H., Sang, P.B., Oh, B.H., Varshney, U., and Woo, E.J. (2019). Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision. Nat Chem Biol *15*, 607-614.
5. Alawneh, A.M., Qi, D., Yonesaki, T., and Otsuka, Y. (2016). An ADP-ribosyltransferase Alt of bacteriophage T4 negatively regulates the Escherichia coli MazF toxin of a toxin-antitoxin module. Mol Microbiol *99*, 188-198.
6. Alexeeva, M., Moen, M.N., Grosvik, K., Tesfahun, A.N., Xu, X.M., Muruzabal-Lecumberri, I., Olsen, K.M., Rasmussen, A., Ruoff, P., Kirpekar, F.*, et al.* (2019). Excision of uracil from DNA by hSMUG1 includes strand incision and processing. Nucleic Acids Res *47*, 779-793.
7. Amundsen, S.K., Sharp, J.W., and Smith, G.R. (2016). RecBCD Enzyme "Chi Recognition" Mutants Recognize Chi Recombination Hotspots in the Right DNA Context. Genetics *204*, 139-152.
8. Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *513*, 569-573.
9. Anderson, D.G., Churchill, J.J., and Kowalczykowski, S.C. (1997). Chi-activated RecBCD enzyme possesses 5'-->3' nucleolytic activity, but RecBC enzyme does not: evidence suggesting that the alteration induced by Chi is not simply ejection of the RecD subunit. Genes Cells *2*, 117-128.
10. Anderson, D.G., Churchill, J.J., and Kowalczykowski, S.C. (1999). A single mutation, RecB(D1080A,) eliminates RecA protein loading but not Chi recognition by RecBCD enzyme. J Biol Chem *274*, 27139-27144.
11. Anderson, D.G., and Kowalczykowski, S.C. (1997a). The recombination hot spot chi is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. Genes Dev *11*, 571-581.
12. Anderson, D.G., and Kowalczykowski, S.C. (1997b). The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. Cell *90*, 77-86.

13. Aravind, L., and Koonin, E.V. (2000). The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates. Genome Biol *1*, RESEARCH0007.
14. Aravind, L., Zhang, D., de Souza, R.F., Anand, S., and Iyer, L.M. (2015). The natural history of ADP-ribosyltransferases and the ADP-ribosylation system. Curr Top Microbiol Immunol *384*, 3-32.
15. Arnold, D.A., and Kowalczykowski, S.C. (2000). Facilitated loading of RecA protein is essential to recombination by RecBCD enzyme. J Biol Chem *275*, 12261-12265.
16. Atanasiu, C., Su, T.J., Sturrock, S.S., and Dryden, D.T. (2002). Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. Nucleic Acids Res *30*, 3936-3944.
17. Athukoralage, J.S., McMahon, S.A., Zhang, C., Gruschow, S., Graham, S., Krupovic, M., Whitaker, R.J., Gloster, T.M., and White, M.F. (2020). An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. Nature *577*, 572-575.
18. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol *2*, 2006 0008.
19. Bailly, V., and Verly, W.G. (1987). Escherichia coli endonuclease III is not an endonuclease but a beta-elimination catalyst. Biochem J *242*, 565-572.
20. Bair, C.L., and Black, L.W. (2007). A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. J Mol Biol *366*, 768-778.
21. Bair, C.L., Rifat, D., and Black, L.W. (2007). Exclusion of glucosyl-hydroxymethylcytosine DNA containing bacteriophages is overcome by the injected protein inhibitor IPI*. J Mol Biol *366*, 779-789.
22. Banh, D.V., Roberts, C.G., Morales-Amador, A., Berryhill, B.A., Chaudhry, W., Levin, B.R., Brady, S.F., and Marraffini, L.A. (2023). Bacterial cGAS senses a viral RNA to initiate immunity. Nature.
23. Bardgett, R.D., and van der Putten, W.H. (2014). Belowground biodiversity and ecosystem functioning. Nature *515*, 505-511.
24. Bari, S.M.N., Chou-Zheng, L., Howell, O., Hossain, M., Hill, C.M., Boyle, T.A., Cater, K., Dandu, V.S., Thomas, A., Aslan, B.*, et al.* (2022). A unique mode of nucleic acid immunity performed by a multifunctional bacterial enzyme. Cell Host Microbe *30*, 570-582 e577.
25. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.
26. Barrett, T.E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., and Pearl, L.H. (1998). Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. Cell *92*, 117-129.

27. Baute, J., and Depicker, A. (2008). Base excision repair and its role in maintaining genome stability. Crit Rev Biochem Mol Biol *43*, 239-276.
28. Behme, M.T., Lilley, G.D., and Ebisuzaki, K. (1976). Postinfection control by bacteriophage T4 of Escherichia coli recBC nuclease activity. J Virol *18*, 20-25.
29. Bernheim, A., and Sorek, R. (2020). The pan-immune system of bacteria: antiviral defence as a community resource. Nat Rev Microbiol *18*, 113-119.
30. Bianco, P.R., and Kowalczykowski, S.C. (1997). The recombination hotspot Chi is recognized by the translocating RecBCD enzyme as the single strand of DNA containing the sequence 5'-GCTGGTGG-3'. Proc Natl Acad Sci U S A *94*, 6706-6711.
31. Bickle, T.A., and Kruger, D.H. (1993). Biology of DNA restriction. Microbiol Rev *57*, 434-450.
32. Bikard, D., Hatoum-Aslan, A., Mucida, D., and Marraffini, L.A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. Cell Host Microbe *12*, 177-186.
33. Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res *41*, 7429-7437.
34. Bobay, L.M., Touchon, M., and Rocha, E.P. (2013). Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. PLoS Genet *9*, e1003825.
35. Bobonis, J., Mitosch, K., Mateus, A., Karcher, N., Kritikos, G., Selkrig, J., Zietek, M., Monzon, V., Pfalz, B., Garcia-Santamarina, S.*, et al.* (2022). Bacterial retrons encode phage-defending tripartite toxin-antitoxin systems. Nature *609*, 144-150.
36. Boiteux, S. (1993). Properties and biological functions of the NTH and FPG proteins of Escherichia coli: two DNA glycosylases that repair oxidative damage in DNA. J Photochem Photobiol B *19*, 87-96.
37. Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551-2561.
38. Bondy-Denomy, J., Davidson, A.R., Doudna, J.A., Fineran, P.C., Maxwell, K.L., Moineau, S., Peng, X., Sontheimer, E.J., and Wiedenheft, B. (2018). A Unified Resource for Tracking Anti-CRISPR Names. CRISPR J *1*, 304-305.
39. Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M.F., Hidalgo-Reyes, Y., Wiedenheft, B., Maxwell, K.L., and Davidson, A.R. (2015). Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. Nature *526*, 136-139.
40. Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., and Davidson, A.R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature *493*, 429-432.
41. Borges, A.L., Zhang, J.Y., Rollins, M.F., Osuna, B.A., Wiedenheft, B., and Bondy-Denomy, J. (2018). Bacteriophage Cooperation Suppresses CRISPR-Cas3 and Cas9 Immunity. Cell *174*, 917-925 e910.

42. Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.F., Lio, P., Crescenzi, P., Fani, R., and Fondi, M. (2015). MeDuSa: a multi-draft based scaffolder. Bioinformatics *31*, 2443-2451.
43. Brady, S.F. (2007). Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. Nat Protoc *2*, 1297-1305.
44. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.
45. Brüssow, H., and Hendrix, R.W. (2002). Phage genomics: small is beautiful. Cell *108*, 13-16.
46. Bryson, A.L., Hwang, Y., Sherrill-Mix, S., Wu, G.D., Lewis, J.D., Black, L., Clark, T.A., and Bushman, F.D. (2015). Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9. mBio *6*, e00648.
47. Caliando, B.J., and Voigt, C.A. (2015). Targeted DNA degradation using a CRISPR device stably carried in the host genome. Nat Commun *6*, 6989.
48. Carlson, K., and Wiberg, J.S. (1983). In vivo cleavage of cytosine-containing bacteriophage T4 DNA to genetically distinct, discretely sized fragments. J Virol *48*, 18-30.
49. Chaveroche, M.K., Ghigo, J.M., and d'Enfert, C. (2000). A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans*. Nucleic Acids Res *28*, E97.
50. Chayot, R., Montagne, B., Mazel, D., and Ricchetti, M. (2010). An end-joining repair mechanism in Escherichia coli. Proc Natl Acad Sci U S A *107*, 2141-2146.
51. Chedin, F., Handa, N., Dillingham, M.S., and Kowalczykowski, S.C. (2006). The AddAB helicase/nuclease forms a stable complex with its cognate chi sequence during translocation. J Biol Chem *281*, 18610-18617.
52. Chedin, F., Noirot, P., Biaudet, V., and Ehrlich, S.D. (1998). A five-nucleotide sequence protects DNA from exonucleolytic degradation by AddAB, the RecBCD analogue of Bacillus subtilis. Mol Microbiol *29*, 1369-1377.
53. Chiu, C.S., Tomich, P.K., and Greenberg, G.R. (1976). Simultaneous initiation of synthesis of bacteriophage T4 DNA and of deoxyribonucleotides. Proc Natl Acad Sci U S A *73*, 757-761.
54. Chowdhury, S., Carter, J., Rollins, M.F., Golden, S.M., Jackson, R.N., Hoffmann, C., Nosaka, L., Bondy-Denomy, J., Maxwell, K.L., Davidson, A.R.*, et al.* (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. Cell *169*, 47-57 e11.
55. Churchill, J.J., Anderson, D.G., and Kowalczykowski, S.C. (1999). The RecBC enzyme loads RecA protein onto ssDNA asymmetrically and independently of chi, resulting in constitutive recombination activation. Genes Dev *13*, 901-911.
56. Chylinski, K., Le Rhun, A., and Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol *10*, 726-737.

57. Clement, J.M., Lepouce, E., Marchal, C., and Hofnung, M. (1983). Genetic study of a membrane protein: DNA sequence alterations due to 17 *lamB* point mutations affecting adsorption of phage lambda. EMBO J *2*, 77-80.
58. Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G., and Sorek, R. (2019). Cyclic GMP-AMP signalling protects bacteria against viral infection. Nature *574*, 691-695.
59. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science *339*, 819-823.
60. Connelly, J.C., Kirkham, L.A., and Leach, D.R. (1998). The SbcCD nuclease of Escherichia coli is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. Proc Natl Acad Sci U S A *95*, 7969-7974.
61. Court, R., Cook, N., Saikrishnan, K., and Wigley, D. (2007). The crystal structure of lambda-Gam protein suggests a model for RecBCD inhibition. J Mol Biol *371*, 25-33.
62. Cox, M.M., Goodman, M.F., Keck, J.L., van Oijen, A., Lovett, S.T., and Robinson, A. (2023). Generation and Repair of Postreplication Gaps in Escherichia coli. Microbiol Mol Biol Rev *87*, e0007822.
63. Cromie, G.A., Connelly, J.C., and Leach, D.R. (2001). Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. Mol Cell *8*, 1163-1174.
64. Cui, L., and Bikard, D. (2016). Consequences of Cas9 cleavage in the chromosome of Escherichia coli. Nucleic Acids Res *44*, 4243-4251.
65. Cui, N., Zhang, J.T., Liu, Y., Liu, Y., Liu, X.Y., Wang, C., Huang, H., and Jia, N. (2023). Type IV-A CRISPR-Csf complex: Assembly, dsDNA targeting, and CasDinG recruitment. Mol Cell *83*, 2493-2508 e2495.
66. d'Herelle, F. (1917). Sur un microbe invisible antagoniste des bacilles dysentériques. CR Acad Sci Paris *165*, 373-375.
67. Dai, N., Bitinaite, J., Chin, H.G., Pradhan, S., and Correa, I.R., Jr. (2013). Evaluation of UDP-GlcN derivatives for selective labeling of 5-(hydroxymethyl)cytosine. ChemBioChem *14*, 2144-2152.
68. Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA *97*, 6640-6645.
69. Davidson, A.R., Lu, W.T., Stanley, S.Y., Wang, J., Mejdani, M., Trost, C.N., Hicks, B.T., Lee, J., and Sontheimer, E.J. (2020). Anti-CRISPRs: Protein Inhibitors of CRISPR-Cas Systems. Annu Rev Biochem *89*, 309-332.
70. de Waard, A., Ubbink, T.E., and Beukman, W. (1967). On the specificity of bacteriophage-induced hydroxymethylcytosine glucosyltransferases. II. Specificities of hydroxymethylcytosine alphaand beta-glucosyltransferases induced by bacteriophage T4. Eur J Biochem *2*, 303-308.
71. Deep, A., Gu, Y., Gao, Y.Q., Ego, K.M., Herzik, M.A., Jr., Zhou, H., and Corbett, K.D. (2022). The SMC-family Wadjet complex protects bacteria from

plasmid transformation by recognition and cleavage of closed-circular DNA. Mol Cell *82*, 4145-4159 e4147.
72. Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-Gonzalez, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K., and Fierer, N. (2018). A global atlas of the dominant bacteria found in soil. Science *359*, 320-325.
73. Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.
74. Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol *190*, 1390-1400.
75. Dianov, G., and Lindahl, T. (1994). Reconstitution of the DNA base excision-repair pathway. Curr Biol *4*, 1069-1076.
76. Dianov, G., Sedgwick, B., Daly, G., Olsson, M., Lovett, S., and Lindahl, T. (1994). Release of 5'-terminal deoxyribose-phosphate residues from incised abasic sites in DNA by the Escherichia coli RecJ protein. Nucleic Acids Res *22*, 993-998.
77. Dila, D., Sutherland, E., Moran, L., Slatko, B., and Raleigh, E.A. (1990). Genetic and sequence organization of the mcrBC locus of Escherichia coli K-12. J Bacteriol *172*, 4888-4900.
78. Dillingham, M.S., and Kowalczykowski, S.C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiol Mol Biol Rev *72*, 642-671.
79. Dillingham, M.S., Spies, M., and Kowalczykowski, S.C. (2003). RecBCD enzyme is a bipolar DNA helicase. Nature *423*, 893-897.
80. Dion, M.B., Oechslin, F., and Moineau, S. (2020). Phage diversity, genomics and phylogeny. Nat Rev Microbiol *18*, 125-138.
81. Dixon, D.A., and Kowalczykowski, S.C. (1991). Homologous pairing in vitro stimulated by the recombination hotspot, Chi. Cell *66*, 361-371.
82. Dixon, D.A., and Kowalczykowski, S.C. (1993). The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the E. coli RecBCD enzyme. Cell *73*, 87-96.
83. Dixon, D.A., and Kowalczykowski, S.C. (1995). Role of the Escherichia coli recombination hotspot, chi, in RecABCD-dependent homologous pairing. J Biol Chem *270*, 16360-16370.
84. Dong, Guo, M., Wang, S., Zhu, Y., Wang, S., Xiong, Z., Yang, J., Xu, Z., and Huang, Z. (2017). Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. Nature *546*, 436-439.
85. Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., and Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. Science *359*.
86. Drake, J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci USA *88*, 7160-7164.

87. Drivdahl, R.H., and Kutter, E.M. (1990). Inhibition of transcription of cytosine-containing DNA in vitro by the alc gene product of bacteriophage T4. J Bacteriol *172*, 2716-2727.
88. Dutta, S., Chowdhury, G., and Gates, K.S. (2007). Interstrand cross-links generated by abasic sites in duplex DNA. J Am Chem Soc *129*, 1852-1853.
89. Echolas, H., and Gingery, R. (1968). Mutants of bacteriophage lambda defective in vegetative genetic recombination. J Mol Biol *34*, 239-249.
90. Elmore, J.R., Sheppard, N.F., Ramia, N., Deighan, T., Li, H., Terns, R.M., and Terns, M.P. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. Genes Dev *30*, 447-459.
91. Enquist, L.W., and Skalka, A. (1973). Replication of bacteriophage lambda DNA dependent on the function of host and viral genes. I. Interaction of red, gam and rec. J Mol Biol *75*, 185-212.
92. Feng, Z., Kallifidas, D., and Brady, S.F. (2011). Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. Proc Natl Acad Sci U S A *108*, 12629-12634.
93. Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398-402.
94. Fillol-Salom, A., Rostol, J.T., Ojiogu, A.D., Chen, J., Douce, G., Humphrey, S., and Penades, J.R. (2022). Bacteriophages benefit from mobilizing pathogenicity islands encoding immune systems against competitors. Cell *185*, 3248-3262 e3220.
95. Fortini, P., and Dogliotti, E. (2007). Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. DNA Repair (Amst) *6*, 398-409.
96. Fu, B.X., Wainberg, M., Kundaje, A., and Fire, A.Z. (2017). High-Throughput Characterization of Cascade type I-E CRISPR Guide Efficacy Reveals Unexpected PAM Diversity and Target Sequence Preferences. Genetics *206*, 1727-1738.
97. Fukuyo, M., Nakano, T., Zhang, Y., Furuta, Y., Ishikawa, K., Watanabe-Matsui, M., Yano, H., Hamakawa, T., Ide, H., and Kobayashi, I. (2015). Restriction-modification system with methyl-inhibited base excision and abasic-site cleavage activities. Nucleic Acids Res *43*, 2841-2852.
98. Galletto, R., Amitani, I., Baskin, R.J., and Kowalczykowski, S.C. (2006). Direct observation of individual RecA filaments assembling on single DNA molecules. Nature *443*, 875-878.
99. Gao, L., Altae-Tran, H., Bohning, F., Makarova, K.S., Segel, M., Schmid-Burgk, J.L., Koob, J., Wolf, Y.I., Koonin, E.V., and Zhang, F. (2020). Diverse enzymatic activities mediate antiviral immunity in prokaryotes. Science *369*, 1077-1084.

100. Gao, L.A., Wilkinson, M.E., Strecker, J., Makarova, K.S., Macrae, R.K., Koonin, E.V., and Zhang, F. (2022). Prokaryotic innate immunity through pattern recognition of conserved viral proteins. Science *377*, eabm4096.
101. Garb, J., Lopatina, A., Bernheim, A., Zaremba, M., Siksnys, V., Melamed, S., Leavitt, A., Millman, A., Amitai, G., and Sorek, R. (2022). Multiple phage resistance systems inhibit infection via SIR2-dependent NAD⁺ depletion. Nat Microbiol *7*, 1849-1856.
102. Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67-71.
103. Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci USA *109*, E2579-2586.
104. Georgopoulos, C.P., and Revel, H.R. (1971). Studies with glucosyl transferase mutants of the T-even bacteriophages. Virology *44*, 271-285.
105. Georjon, H., and Bernheim, A. (2023). The highly diverse antiphage defence systems of bacteria. Nat Rev Microbiol *21*, 686-700.
106. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.
107. Goldberg, G.W., Jiang, W., Bikard, D., and Marraffini, L.A. (2014). Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. Nature *514*, 633-637.
108. Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G., and Sorek, R. (2015). BREX is a novel phage resistance system widespread in microbial genomes. EMBO J *34*, 169-183.
109. Gong, C., Bongiorno, P., Martins, A., Stephanou, N.C., Zhu, H., Shuman, S., and Glickman, M.S. (2005). Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. Nat Struct Mol Biol *12*, 304-312.
110. Gordeeva, J., Morozova, N., Sierro, N., Isaev, A., Sinkunas, T., Tsvetkova, K., Matlashov, M., Truncaite, L., Morgan, R.D., Ivanov, N.V.*, et al.* (2018). BREX system of Escherichia coli distinguishes self from non-self by methylation of a specific DNA site. Nucleic Acids Res.
111. Guegler, C.K., and Laub, M.T. (2021). Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection. Mol Cell *81*, 2361-2373 e2369.
112. Guo, T.W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A., Eng, E.T., Raczkowski, A.M., Fox, T., Earl, L.A.*, et al.* (2017). Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. Cell *171*, 414-426 e412.

113. Guyer, M.S., Reed, R.R., Steitz, J.A., and Low, K.B. (1981). Identification of a sex-factor-affinity site in E. coli as gamma delta. Cold Spring Harb Symp Quant Biol *45 Pt 1*, 135-140.
114. Hale, C., Kleppe, K., Terns, R.M., and Terns, M.P. (2008). Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. RNA *14*, 2572-2579.
115. Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945-956.
116. Hampton, H.G., Watson, B.N.J., and Fineran, P.C. (2020). The arms race between bacteria and their phage foes. Nature *577*, 327-336.
117. Handa, N., Yang, L., Dillingham, M.S., Kobayashi, I., Wigley, D.B., and Kowalczykowski, S.C. (2012). Molecular determinants responsible for recognition of the single-stranded DNA regulatory sequence, chi, by RecBCD enzyme. Proc Natl Acad Sci U S A *109*, 8901-8906.
118. Harrington, L.B., Doxzen, K.W., Ma, E., Liu, J.J., Knott, G.J., Edraki, A., Garcia, B., Amrani, N., Chen, J.S., Cofsky, J.C.*, et al.* (2017). A Broad-Spectrum Inhibitor of CRISPR-Cas9. Cell *170*, 1224-1233 e1215.
119. Harvey, H., Bondy-Denomy, J., Marquis, H., Sztanko, K.M., Davidson, A.R., and Burrows, L.L. (2018). Pseudomonas aeruginosa defends against phages through type IV pilus glycosylation. Nat Microbiol *3*, 47-52.
120. Hedglin, M., and O'Brien, P.J. (2010). Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein. ACS Chem Biol *5*, 427-436.
121. Heitman, J., and Model, P. (1987). Site-specific methylases induce the SOS DNA repair response in Escherichia coli. J Bacteriol *169*, 3243-3250.
122. Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature *519*, 199-202.
123. Heler, R., Wright, A.V., Vucelja, M., Doudna, J.A., and Marraffini, L.A. (2019). Spacer Acquisition Rates Determine the Immunological Diversity of the Type II CRISPR-Cas Immune Response. Cell Host Microbe *25*, 242-249 e243.
124. Henderson, D., and Weil, J. (1975). Recombination-deficient deletions in bacteriophage lambda and their interaction with chi mutations. Genetics *79*, 143-174.
125. Henrikus, S.S., van Oijen, A.M., and Robinson, A. (2018). Specialised DNA polymerases in Escherichia coli: roles within multiple pathways. Curr Genet *64*, 1189-1196.
126. Herskowitz, I., and Hagen, D. (1980). The lysis-lysogeny decision of phage lambda: explicit programming and responsiveness. Annu Rev Genet *14*, 399-445.
127. Heyer, W.D. (2004). Damage signaling: RecQ sends an SOS to you. Curr Biol *14*, R895-897.
128. Hirano, N., Ohshima, H., and Takahashi, H. (2006). Biochemical analysis of the substrate specificity and sequence preference of endonuclease IV from

bacteriophage T4, a dC-specific endonuclease implicated in restriction of dC-substituted T4 DNA synthesis. Nucleic Acids Res *34*, 4743-4751.
129. Ho, C.H., Wang, H.C., Ko, T.P., Chang, Y.C., and Wang, A.H. (2014). The T4 phage DNA mimic protein Arn inhibits the DNA binding activity of the bacterial histone-like protein H-NS. J Biol Chem *289*, 27046-27054.
130. Hobbs, S.J., Wein, T., Lu, A., Morehouse, B.R., Schnabel, J., Leavitt, A., Yirmiya, E., Sorek, R., and Kranzusch, P.J. (2022). Phage anti-CBASS and anti-Pycsar nucleases subvert bacterial immunity. Nature *605*, 522-526.
131. Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proc Natl Acad Sci U S A *111*, 6618-6623.
132. Holm, L. (2022). Dali server: structural unification of protein families. Nucleic Acids Res *50*, W210-W215.
133. Hoskisson, P.A., Sumby, P., and Smith, M.C.M. (2015). The phage growth limitation system in Streptomyces coelicolor A(3)2 is a toxin/antitoxin system, comprising enzymes with DNA methyltransferase, protein kinase and ATPase activity. Virology *477*, 100-109.
134. Hossain, A.A., McGinn, J., Meeske, A.J., Modell, J.W., and Marraffini, L.A. (2021). Viral recombination systems limit CRISPR-Cas targeting through the generation of escape mutations. Cell Host Microbe.
135. Hsueh, B.Y., Severin, G.B., Elg, C.A., Waldron, E.J., Kant, A., Wessel, A.J., Dover, J.A., Rhoades, C.R., Ridenhour, B.J., Parent, K.N.*, et al.* (2022). Phage defence by deaminase-mediated depletion of deoxynucleotides in bacteria. Nat Microbiol *7*, 1210-1220.
136. Huiting, E., Cao, X., Ren, J., Athukoralage, J.S., Luo, Z., Silas, S., An, N., Carion, H., Zhou, Y., Fraser, J.S.*, et al.* (2023). Bacteriophages inhibit and evade cGAS-like immune function in bacteria. Cell *186*, 864-876.
137. Huo, Y., Nam, K.H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M.D., Jr., Zhou, S., Rajashankar, K., Kurinov, I.*, et al.* (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. Nat Struct Mol Biol *21*, 771-777.
138. Hutinet, G., Kot, W., Cui, L., Hillebrand, R., Balamkundu, S., Gnanakalai, S., Neelakandan, R., Carstens, A.B., Fa Lui, C., Tremblay, D.*, et al.* (2019). 7-Deazaguanine modifications protect phage DNA from host restriction systems. Nat Commun *10*, 5442.
139. Isaev, A., Drobiazko, A., Sierro, N., Gordeeva, J., Yosef, I., Qimron, U., Ivanov, N.V., and Severinov, K. (2020). Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence. Nucleic Acids Res *48*, 5397-5406.
140. Ishikawa, H., Ma, Z., and Barber, G.N. (2009). STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. Nature *461*, 788-792.
141. Ishikawa, K., Watanabe, M., Kuroita, T., Uchiyama, I., Bujnicki, J.M., Kawakami, B., Tanokura, M., and Kobayashi, I. (2005). Discovery of a novel

restriction endonuclease by genome comparison and application of a wheat-germ-based cell-free translation assay: PabI (5'-GTA/C) from the hyperthermophilic archaeon Pyrococcus abyssi. Nucleic Acids Res *33*, e112.
142. Jacobs, A.L., and Schar, P. (2012). DNA glycosylases: in DNA repair and beyond. Chromosoma *121*, 1-20.
143. Jakociune, D., and Moodley, A. (2018). A Rapid Bacteriophage DNA Extraction Method. Methods Protoc *1*.
144. Janeway, C.A., Jr. (1989). Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harb Symp Quant Biol *54 Pt 1*, 1-13.
145. Jiang, F., and Doudna, J.A. (2017). CRISPR-Cas9 Structures and Mechanisms. Annu Rev Biophys *46*, 505-529.
146. Jiang, F., Liu, J.J., Osuna, B.A., Xu, M., Berry, J.D., Rauch, B.J., Nogales, E., Bondy-Denomy, J., and Doudna, J.A. (2019). Temperature-Responsive Competitive Inhibition of CRISPR-Cas9. Mol Cell *73*, 601-610 e605.
147. Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J.A. (2015). STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. Science *348*, 1477-1481.
148. Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol *31*, 233-239.
149. Jiang, W., Samai, P., and Marraffini, L.A. (2016). Degradation of phage transcripts by CRISPR-associated RNases enables type III CRISPR-Cas immunity. Cell *164*, 710-721.
150. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.
151. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. eLife *2*, e00471.
152. Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S.*, et al.* (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science *343*, 1247997.
153. Johnson, A.G., Wein, T., Mayer, M.L., Duncan-Lowey, B., Yirmiya, E., Oppenheimer-Shaanan, Y., Amitai, G., Sorek, R., and Kranzusch, P.J. (2022). Bacterial gasdermins reveal an ancient mechanism of cell death. Science *375*, 221-225.
154. Jones, J.D., Vance, R.E., and Dangl, J.L. (2016). Intracellular innate immune surveillance devices in plants and animals. Science *354*.
155. Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R.*, et al.* (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol *18*, 529-536.

156. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A.*, et al.* (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589.
157. Karakousis, G., Ye, N., Li, Z., Chiu, S.K., Reddy, G., and Radding, C.M. (1998). The beta protein of phage lambda binds preferentially to an intermediate in DNA renaturation. J Mol Biol *276*, 721-731.
158. Karam, J.D., and Miller, E.S. (2010). Bacteriophage T4 and its relatives. Virol J *7*, 293.
159. Karu, A.E., Sakaki, Y., Echols, H., and Linn, S. (1975). The gamma protein specified by bacteriophage gamma. Structure and inhibitory activity for the recBC enzyme of Escherichia coli. J Biol Chem *250*, 7377-7387.
160. Kawai, A., Higuchi, S., Tsunoda, M., Nakamura, K.T., Yamagata, Y., and Miyamoto, S. (2015). Crystal structure of family 4 uracil-DNA glycosylase from Sulfolobus tokodaii and a function of tyrosine 170 in DNA binding. FEBS Lett *589*, 2675-2682.
161. Kazlauskiene, M., Kostiuk, G., Venclovas, C., Tamulaitis, G., and Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science *357*, 605-609.
162. Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, C., and Siksnys, V. (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. Mol Cell *62*, 295-306.
163. Khudyakov, I.Y., Kirnos, M.D., Alexandrushkina, N.I., and Vanyushin, B.F. (1978). Cyanophage S-2L contains DNA with 2,6-diaminopurine substituted for adenine. Virology *88*, 8-18.
164. Kirnos, M.D., Khudyakov, I.Y., Alexandrushkina, N.I., and Vanyushin, B.F. (1977). 2-aminoadenine is an adenine substituting for a base in S-2L cyanophage DNA. Nature *270*, 369-370.
165. Kmiec, E., and Holloman, W.K. (1981). Beta protein of bacteriophage lambda promotes renaturation of DNA. J Biol Chem *256*, 12636-12639.
166. Knott, G.J., Thornton, B.W., Lobba, M.J., Liu, J.J., Al-Shayeb, B., Watters, K.E., and Doudna, J.A. (2019). Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. Nat Struct Mol Biol *26*, 315-321.
167. Ko, C.C., and Hatfull, G.F. (2018). Mycobacteriophage Fruitloop gp52 inactivates Wag31 (DivIVA) to prevent heterotypic superinfection. Mol Microbiol *108*, 443-460.
168. Koga, M., Otsuka, Y., Lemire, S., and Yonesaki, T. (2011). Escherichia coli rnlA and rnlB compose a novel toxin-antitoxin system. Genetics *187*, 123-130.
169. Korn, A.M., Hillhouse, A.E., Sun, L., and Gill, J.J. (2021). Comparative Genomics of Three Novel Jumbo Bacteriophages Infecting Staphylococcus aureus. J Virol *95*, e0239120.
170. Koths, K., and Dressler, D. (1978). Analysis of the phiX DNA replication cycle by electron microscopy. Proc Natl Acad Sci U S A *75*, 605-609.
171. Kowalczykowski, S.C. (2000). Initiation of genetic recombination and recombination-dependent replication. Trends Biochem Sci *25*, 156-165.

172. Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D., and Rehrauer, W.M. (1994). Biochemistry of homologous recombination in Escherichia coli. Microbiol Rev *58*, 401-465.
173. Krokan, H.E., and Bjoras, M. (2013). Base excision repair. Cold Spring Harb Perspect Biol *5*, a012583.
174. Kropinski, A.M., Turner, D., Nash, J.H.E., Ackermann, H.W., Lingohr, E.J., Warren, R.A., Ehrlich, K.C., and Ehrlich, M. (2018). The Sequence of Two Bacteriophages with Hypermodified Bases Reveals Novel Phage-Host Interactions. Viruses *10*.
175. Kruger, D.H., and Bickle, T.A. (1983). Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. Microbiol Rev *47*, 345-360.
176. Kuhn, H., Protozanova, E., and Demidov, V.V. (2002). Monitoring of single nicks in duplex DNA by gel electrophoretic mobility-shift assay. Electrophoresis *23*, 2384-2387.
177. Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. Annu Rev Biochem *74*, 681-710.
178. Kuno, S., and Lehman, I.R. (1962). Gentiobiose, a constituent of deoxyribonucleic acid from coliphage T6. J Biol Chem *237*, 1266-1270.
179. Kuzminov, A. (1995). Collapse and repair of replication forks in Escherichia coli. Mol Microbiol *16*, 373-384.
180. Kuzminov, A. (1999). Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. Microbiol Mol Biol Rev *63*, 751-813.
181. Lam, S.T., Stahl, M.M., McMilin, K.D., and Stahl, F.W. (1974). Rec-mediated recombinational hot spot activity in bacteriophage lambda. II. A mutation which causes hot spot activity. Genetics *77*, 425-433.
182. Lamm, N., Wang, Y., Mathews, C.K., and Ruger, W. (1988). Deoxycytidylate hydroxymethylase gene of bacteriophage T4. Nucleotide sequence determination and over-expression of the gene. Eur J Biochem *172*, 553-563.
183. Landsberger, M., Gandon, S., Meaden, S., Rollie, C., Chevallereau, A., Chabas, H., Buckling, A., Westra, E.R., and van Houte, S. (2018). Anti-CRISPR Phages Cooperate to Overcome CRISPR-Cas Immunity. Cell *174*, 908-916 e912.
184. Lau, R.K., Ye, Q., Birkholz, E.A., Berg, K.R., Patel, L., Mathews, I.T., Watrous, J.D., Ego, K., Whiteley, A.T., Lowey, B.*, et al.* (2020). Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. Mol Cell *77*, 723-733 e726.
185. Le, T.B., Imakaev, M.V., Mirny, L.A., and Laub, M.T. (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. Science *342*, 731-734.
186. Leavitt, A., Yirmiya, E., Amitai, G., Lu, A., Garb, J., Herbst, E., Morehouse, B.R., Hobbs, S.J., Antine, S.P., Sun, Z.J.*, et al.* (2022). Viruses inhibit TIR gcADPR signalling to overcome bacterial defence. Nature *611*, 326-331.

187. Lee, J., Mir, A., Edraki, A., Garcia, B., Amrani, N., Lou, H.E., Gainetdinov, I., Pawluk, A., Ibraheim, R., Gao, X.D.*, et al.* (2018). Potent Cas9 Inhibition in Bacterial and Human Cells by AcrIIC4 and AcrIIC5 Anti-CRISPR Proteins. MBio *9*.
188. Lehman, I.R., and Pratt, E.A. (1960). On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. J Biol Chem *235*, 3254-3259.
189. Lenhart, J.S., Schroeder, J.W., Walsh, B.W., and Simmons, L.A. (2012). DNA repair and genome maintenance in Bacillus subtilis. Microbiol Mol Biol Rev *76*, 530-564.
190. LeRoux, M., and Laub, M.T. (2022). Toxin-Antitoxin Systems as Phage Defense Elements. Annu Rev Microbiol *76*, 21-43.
191. LeRoux, M., Srikant, S., Teodoro, G.I.C., Zhang, T., Littlehale, M.L., Doron, S., Badiee, M., Leung, A.K.L., Sorek, R., and Laub, M.T. (2022). The DarTG toxin-antitoxin system provides phage defence by ADP-ribosylating viral DNA. Nat Microbiol *7*, 1028-1040.
192. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res *49*, W293-W296.
193. Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., and Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature *520*, 505-510.
194. Lhomme, J., Constant, J.F., and Demeunynck, M. (1999). Abasic DNA structure, reactivity, and recognition. Biopolymers *52*, 65-83.
195. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.
196. Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709-715.
197. Liow, L.H., Van Valen, L., and Stenseth, N.C. (2011). Red Queen: from populations to taxa and communities. Trends Ecol Evol *26*, 349-358.
198. Little, J.W. (1967). An exonuclease induced by bacteriophage lambda. II. Nature of the enzymatic reaction. J Biol Chem *242*, 679-686.
199. Little, J.W., Lehman, I.R., and Kaiser, A.D. (1967). An exonuclease induced by bacteriophage lambda. I. Preparation of the crystalline enzyme. J Biol Chem *242*, 672-678.
200. Liu, H.W., Roisne-Hamelin, F., Beckert, B., Li, Y., Myasnikov, A., and Gruber, S. (2022). DNA-measuring Wadjet SMC ATPases restrict smaller circular plasmids by DNA cleavage. Mol Cell *82*, 4727-4740 e4726.
201. Liu, L., Yin, M., Wang, M., and Wang, Y. (2019a). Phage AcrIIA2 DNA Mimicry: Structural Basis of the CRISPR and Anti-CRISPR Arms Race. Mol Cell *73*, 611-620 e613.
202. Liu, M., Li, C.C., Luo, X., Ma, F., and Zhang, C.Y. (2020a). 5-Hydroxymethylcytosine Glucosylation-Triggered Helicase-Dependent Amplification-Based Fluorescent Biosensor for Sensitive Detection of beta-

Glucosyltransferase with Zero Background Signal. Anal Chem *92*, 16307-16313.
203. Liu, Y., Dai, L., Dong, J., Chen, C., Zhu, J., Rao, V.B., and Tao, P. (2020b). Covalent Modifications of the Bacteriophage Genome Confer a Degree of Resistance to Bacterial CRISPR Systems. J Virol *94*.
204. Liu, Z.J., Martinez Cuesta, S., van Delft, P., and Balasubramanian, S. (2019b). Sequencing abasic sites in DNA at single-nucleotide resolution. Nat Chem *11*, 629-637.
205. Loenen, W.A., Dryden, D.T., Raleigh, E.A., and Wilson, G.G. (2014a). Type I restriction enzymes and their relatives. Nucleic Acids Res *42*, 20-44.
206. Loenen, W.A., Dryden, D.T., Raleigh, E.A., Wilson, G.G., and Murray, N.E. (2014b). Highlights of the DNA cutters: a short history of the restriction enzymes. Nucleic Acids Res *42*, 3-19.
207. Loenen, W.A., and Raleigh, E.A. (2014). The other face of restriction: modification-dependent enzymes. Nucleic Acids Res *42*, 56-69.
208. Lopatina, A., Tal, N., and Sorek, R. (2020). Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. Annu Rev Virol *7*, 371-384.
209. Lopes, A., Amarir-Bouhram, J., Faure, G., Petit, M.A., and Guerois, R. (2010). Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. Nucleic Acids Res *38*, 3952-3962.
210. Lowey, B., Whiteley, A.T., Keszei, A.F.A., Morehouse, B.R., Mathews, I.T., Antine, S.P., Cabrera, V.J., Kashin, D., Niemann, P., Jain, M.*, et al.* (2020). CBASS Immunity Uses CARF-Related Effectors to Sense 3'-5'- and 2'-5'-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. Cell *182*, 38-49 e17.
211. Lunt, M.R., and Newton, E.A. (1965). Glucosylated Nucleotide Sequences from T-Even Bacteriophage Deoxyribonucleic Acids. Biochem J *95*, 717-723.
212. Luria, S.E., and Human, M.L. (1952). A nonhereditary, host-induced variation of bacterial viruses. J Bacteriol *64*, 557-569.
213. Lusetti, S.L., and Cox, M.M. (2002). The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. Annu Rev Biochem *71*, 71-100.
214. Machnicka, M.A., Kaminska, K.H., Dunin-Horkawicz, S., and Bujnicki, J.M. (2015). Phylogenomics and sequence-structure-function relationships in the GmrSD family of Type IV restriction enzymes. BMC Bioinformatics *16*, 336.
215. Maffei, E., Shaidullina, A., Burkolter, M., Heyer, Y., Estermann, F., Druelle, V., Sauer, P., Willi, L., Michaelis, S., Hilbi, H.*, et al.* (2021). Systematic exploration of Escherichia coli phage-host interactions with the BASEL phage collection. PLoS Biol *19*, e3001424.
216. Maguin, P., Varble, A., Modell, J.W., and Marraffini, L.A. (2022). Cleavage of viral DNA by restriction endonucleases stimulates the type II CRISPR-Cas immune response. Mol Cell *82*, 907-919.
217. Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P.*, et al.*

(2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol *18*, 67-83.
218. Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2013). Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Res *41*, 4360-4377.
219. Makarova, K.S., Wolf, Y.I., Snir, S., and Koonin, E.V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. J Bacteriol *193*, 6039-6056.
220. Marcy, Y., Ouverney, C., Bik, E.M., Losekann, T., Ivanova, N., Martin, H.G., Szeto, E., Platt, D., Hugenholtz, P., Relman, D.A.*, et al.* (2007). Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proc Natl Acad Sci U S A *104*, 11889-11894.
221. Marino, N.D., Zhang, J.Y., Borges, A.L., Sousa, A.A., Leon, L.M., Rauch, B.J., Walton, R.T., Berry, J.D., Joung, J.K., Kleinstiver, B.P.*, et al.* (2018). Discovery of widespread type I and type V CRISPR-Cas inhibitors. Science *362*, 240-242.
222. Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55-61.
223. Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843-1845.
224. Martinsohn, J.T., Radman, M., and Petit, M.A. (2008). The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. PLoS Genet *4*, e1000065.
225. Marty, M.T., Baldwin, A.J., Marklund, E.G., Hochberg, G.K., Benesch, J.L., and Robinson, C.V. (2015). Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. Anal Chem *87*, 4370-4376.
226. Masaoka, A., Matsubara, M., Hasegawa, R., Tanaka, T., Kurisu, S., Terato, H., Ohyama, Y., Karino, N., Matsuda, A., and Ide, H. (2003). Mammalian 5-formyluracil-DNA glycosylase. 2. Role of SMUG1 uracil-DNA glycosylase in repair of 5-formyluracil and other oxidized and deaminated base lesions. Biochemistry *42*, 5003-5012.
227. Maslowska, K.H., Makiela-Dzbenska, K., and Fijalkowska, I.J. (2019). The SOS system: A complex and tightly regulated response to DNA damage. Environ Mol Mutagen *60*, 368-384.
228. McGinn, J., and Marraffini, L.A. (2016). CRISPR-Cas systems optimize their immune response by specifying the site of spacer integration. Mol Cell *64*, 616-623.
229. McGinn, J., and Marraffini, L.A. (2019). Molecular mechanisms of CRISPR-Cas spacer acquisition. Nat Rev Microbiol *17*, 7-12.
230. McVey, M., and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet *24*, 529-538.

231. Meeske, A.J., Nakandakari-Higa, S., and Marraffini, L.A. (2019). Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. Nature *570*, 241-245.
232. Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell *151*, 1417-1430.
233. Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Ruger, W. (2003). Bacteriophage T4 genome. Microbiol Mol Biol Rev *67*, 86-156.
234. Millman, A., Bernheim, A., Stokar-Avihail, A., Fedorenko, T., Voichek, M., Leavitt, A., Oppenheimer-Shaanan, Y., and Sorek, R. (2020a). Bacterial Retrons Function In Anti-Phage Defense. Cell *183*, 1551-1561 e1512.
235. Millman, A., Melamed, S., Amitai, G., and Sorek, R. (2020b). Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. Nat Microbiol *5*, 1608-1615.
236. Millman, A., Melamed, S., Leavitt, A., Doron, S., Bernheim, A., Hor, J., Garb, J., Bechon, N., Brandis, A., Lopatina, A*., et al.* (2022). An expanded arsenal of immune systems that protect bacteria from phages. Cell Host Microbe *30*, 1556-1569 e1555.
237. Miranda, A., and Kuzminov, A. (2003). Chromosomal lesion suppression and removal in Escherichia coli via linear DNA degradation. Genetics *163*, 1255-1271.
238. Mirdita, M., Schutze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat Methods *19*, 679-682.
239. Miyazono, K., Furuta, Y., Watanabe-Matsui, M., Miyakawa, T., Ito, T., Kobayashi, I., and Tanokura, M. (2014). A sequence-specific DNA glycosylase mediates restriction-modification in Pyrococcus abyssi. Nat Commun *5*, 3178.
240. Mo, C.Y., Mathai, J., Rostol, J.T., Varble, A., Banh, D.V., and Marraffini, L.A. (2021). Type III-A CRISPR immunity promotes mutagenesis of staphylococci. Nature *592*, 611-615.
241. Modell, J.W., Jiang, W., and Marraffini, L.A. (2017). CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. Nature *544*, 101-104.
242. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol *60*, 174-182.
243. Mol, C.D., Arvai, A.S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H.E., and Tainer, J.A. (1995). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. Cell *80*, 869-878.
244. Mosberg, J.A., Lajoie, M.J., and Church, G.M. (2010). Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. Genetics *186*, 791-799.

245. Mulepati, S., and Bailey, S. (2013). In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. J Biol Chem *288*, 22184-22192.
246. Muniyappa, K., and Radding, C.M. (1986). The homologous recombination system of phage lambda. Pairing activities of beta protein. J Biol Chem *261*, 7472-7478.
247. Murphy, K.C. (1991). Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of Escherichia coli RecBCD enzyme. J Bacteriol *173*, 5808-5821.
248. Murphy, K.C. (1998). Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. J Bacteriol *180*, 2063-2071.
249. Murphy, K.C. (2007). The lambda Gam protein inhibits RecBCD binding to dsDNA ends. J Mol Biol *371*, 19-24.
250. Murphy, K.C. (2016). lambda Recombination and Recombineering. EcoSal Plus *7*.
251. Nascimento, A.L., Souza, A.J., Andrade, P.A.M., Andreote, F.D., Coscione, A.R., Oliveira, F.C., and Regitano, J.B. (2018). Sewage Sludge Microbial Structures and Relations to Their Sources, Treatments, and Chemical Attributes. Front Microbiol *9*, 1462.
252. Niewoehner, O., Garcia-Doval, C., Rostol, J.T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L.A., and Jinek, M. (2017). Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. Nature *548*, 543-548.
253. Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935-949.
254. Nussenzweig, P.M., and Marraffini, L.A. (2020). Molecular Mechanisms of CRISPR-Cas Immunity in Bacteria. Annu Rev Genet *54*, 93-120.
255. Nussenzweig, P.M., McGinn, J., and Marraffini, L.A. (2019). Cas9 Cleavage of Viral Genomes Primes the Acquisition of New Immunological Memories. Cell Host Microbe *26*, 515-526 e516.
256. Ofir, G., Herbst, E., Baroz, M., Cohen, D., Millman, A., Doron, S., Tal, N., Malheiro, D.B.A., Malitsky, S., Amitai, G.*, et al.* (2021). Antiviral activity of bacterial TIR domains via immune signalling molecules. Nature *600*, 116-120.
257. Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., and Sorek, R. (2018). DISARM is a widespread bacterial defence system with broad anti-phage activities. Nat Microbiol *3*, 90-98.
258. Ofir, G., and Sorek, R. (2018). Contemporary Phage Biology: From Classic Models to New Insights. Cell *172*, 1260-1270.
259. Ogawa, T., Yu, X., Shinohara, A., and Egelman, E.H. (1993). Similarity of the yeast RAD51 filament to the bacterial RecA filament. Science *259*, 1896-1899.
260. Olsen, N.S., Nielsen, T.K., Cui, L., Dedon, P., Neve, H., Hansen, L.H., and Kot, W. (2023). A novel Queuovirinae lineage of Pseudomonas aeruginosa phages encode dPreQ0 DNA modifications with a single GA motif that provide

restriction and CRISPR Cas9 protection in vitro. Nucleic Acids Res *51*, 8663-8676.
261. Oppenheim, A.B., Kobiler, O., Stavans, J., Court, D.L., and Adhya, S. (2005). Switches in Bacteriophage Lambda Development. Annu Rev Genet *39*, 409-429.
262. Parikh, S.S., Mol, C.D., and Tainer, J.A. (1997). Base excision repair enzyme family portrait: integrating the structure and chemistry of an entire DNA repair pathway. Structure *5*, 1543-1550.
263. Parikka, K.J., Le Romancer, M., Wauters, N., and Jacquet, S. (2017). Deciphering the virus-to-prokaryote ratio (VPR): insights into virus-host relationships in a variety of ecosystems. Biol Rev Camb Philos Soc *92*, 1081-1100.
264. Passy, S.I., Yu, X., Li, Z., Radding, C.M., and Egelman, E.H. (1999). Rings and filaments of beta protein from bacteriophage lambda suggest a superfamily of recombination proteins. Proc Natl Acad Sci U S A *96*, 4279-4284.
265. Paulin, K.A., Cortez, D., and Eichman, B.F. (2022). The SOS response-associated peptidase (SRAP) domain of YedK catalyzes ring opening of abasic sites and reversal of its DNA-protein cross-link. J Biol Chem *298*, 102307.
266. Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheimer, E.J., Maxwell, K.L.*, et al.* (2016a). Naturally Occurring Off-Switches for CRISPR-Cas9. Cell *167*, 1829-1838 e1829.
267. Pawluk, A., Davidson, A.R., and Maxwell, K.L. (2018). Anti-CRISPR: discovery, mechanism and function. Nat Rev Microbiol *16*, 12-17.
268. Pawluk, A., Shah, M., Mejdani, M., Calmettes, C., Moraes, T.F., Davidson, A.R., and Maxwell, K.L. (2017). Disabling a Type I-E CRISPR-Cas Nuclease with a Bacteriophage-Encoded Anti-CRISPR Protein. MBio *8*.
269. Pawluk, A., Staals, R.H., Taylor, C., Watson, B.N., Saha, S., Fineran, P.C., Maxwell, K.L., and Davidson, A.R. (2016b). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. Nat Microbiol *1*, 16085.
270. Pearl, L.H. (2000). Structure and function in the uracil-DNA glycosylase superfamily. Mutat Res *460*, 165-181.
271. Peng, R., Xu, Y., Zhu, T., Li, N., Qi, J., Chai, Y., Wu, M., Zhang, X., Shi, Y., Wang, P.*, et al.* (2017). Alternate binding modes of anti-CRISPR viral suppressors AcrF1/2 to Csy surveillance complex revealed by cryo-EM structures. Cell Res *27*, 853-864.
272. Petermann, E., Ziegler, M., and Oei, S.L. (2003). ATP-dependent selection between single nucleotide and long patch base excision repair. DNA Repair (Amst) *2*, 1101-1114.
273. Pinilla-Redondo, R., Shehreen, S., Marino, N.D., Fagerlund, R.D., Brown, C.M., Sorensen, S.J., Fineran, P.C., and Bondy-Denomy, J. (2020). Discovery of multiple anti-CRISPRs highlights anti-defense gene clustering in mobile genetic elements. Nat Commun *11*, 5652.

274. Ponticelli, A.S., Schultz, D.W., Taylor, A.F., and Smith, G.R. (1985). Chi-dependent DNA strand cleavage by RecBC enzyme. Cell *41*, 145-151.
275. Porecha, R.H., and Stivers, J.T. (2008). Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. Proc Natl Acad Sci U S A *105*, 10791-10796.
276. Poteete, A.R. (2008). Involvement of DNA replication in phage lambda Red-mediated homologous recombination. Mol Microbiol *68*, 66-74.
277. Price, N.E., Johnson, K.M., Wang, J., Fekry, M.I., Wang, Y., and Gates, K.S. (2014). Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex DNA. J Am Chem Soc *136*, 3483-3490.
278. Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., and Wagner, R. (2010). Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol *75*, 1495-1512.
279. Pulido, R., and Lang, R. (2019). Dual Specificity Phosphatases: From Molecular Mechanisms to Biological Function. Int J Mol Sci *20*.
280. Pyenson, N.C., Gayvert, K., Varble, A., Elemento, O., and Marraffini, L.A. (2017). Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. Cell Host Microbe *22*, 343-353 e343.
281. Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173-1183.
282. Quinones, J.L., and Demple, B. (2016). When DNA repair goes wrong: BER-generated DNA-protein crosslinks to oxidative lesions. DNA Repair (Amst) *44*, 103-109.
283. Ramirez, K.S., Leff, J.W., Barberan, A., Bates, S.T., Betley, J., Crowther, T.W., Kelly, E.F., Oldfield, E.E., Shaw, E.A., Steenbock, C*., et al.* (2014). Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. Proc Biol Sci *281*.
284. Rappe, M.S., and Giovannoni, S.J. (2003). The uncultured microbial majority. Annu Rev Microbiol *57*, 369-394.
285. Rauch, B.J., Silvis, M.R., Hultquist, J.F., Waters, C.S., McGregor, M.J., Krogan, N.J., and Bondy-Denomy, J. (2017). Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. Cell *168*, 150-158 e110.
286. Redding, S., Sternberg, S.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K., Wiedenheft, B., Doudna, J.A., and Greene, E.C. (2015). Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System. Cell *163*, 854-865.
287. Remick, B.C., Gaidt, M.M., and Vance, R.E. (2023). Effector-Triggered Immunity. Annu Rev Immunol *41*, 453-481.
288. Revel, H.R. (1967). Restriction of nonglucosylated T-even bacteriophage: properties of permissive mutants of Escherichia coli B and K12. Virology *31*, 688-701.

289. Rifat, D., Wright, N.T., Varney, K.M., Weber, D.J., and Black, L.W. (2008). Restriction endonuclease inhibitor IPI* of bacteriophage T4: a novel structure for a dedicated target. J Mol Biol *375*, 720-734.
290. Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A.*, et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature *499*, 431-437.
291. Roman, L.J., and Kowalczykowski, S.C. (1989). Formation of heteroduplex DNA promoted by the combined activities of Escherichia coli recA and recBCD proteins. J Biol Chem *264*, 18340-18348.
292. Rostol, J.T., and Marraffini, L. (2019a). (Ph)ighting Phages: How Bacteria Resist Their Parasites. Cell Host Microbe *25*, 184-194.
293. Rostol, J.T., and Marraffini, L.A. (2019b). Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR-Cas immunity. Nat Microbiol *4*, 656-662.
294. Rostol, J.T., Xie, W., Kuryavyi, V., Maguin, P., Kao, K., Froom, R., Patel, D.J., and Marraffini, L.A. (2021). The Card1 nuclease provides defence during type III CRISPR immunity. Nature *590*, 624-629.
295. Rousset, F., Depardieu, F., Miele, S., Dowding, J., Laval, A.L., Lieberman, E., Garry, D., Rocha, E.P.C., Bernheim, A., and Bikard, D. (2022). Phages and their satellites encode hotspots of antiviral systems. Cell Host Microbe *30*, 740-753 e745.
296. Roy, D., Huguet, K.T., Grenier, F., and Burrus, V. (2020). IncC conjugative plasmids and SXT/R391 elements repair double-strand breaks caused by CRISPR-Cas during conjugation. Nucleic Acids Res.
297. Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A., and Marraffini, L.A. (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. Cell *161*, 1164-1174.
298. Samson, J.E., Magadan, A.H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. Nat Rev Microbiol *11*, 675-687.
299. Sancar, A., and Reardon, J.T. (2004). Nucleotide excision repair in E. coli and man. Adv Protein Chem *69*, 43-71.
300. Saporito, S.M., and Cunningham, R.P. (1988). Nucleotide sequence of the nfo gene of Escherichia coli K-12. J Bacteriol *170*, 5141-5145.
301. Saporito, S.M., Smith-White, B.J., and Cunningham, R.P. (1988). Nucleotide sequence of the xth gene of Escherichia coli K-12. J Bacteriol *170*, 4542-4547.
302. Sashital, D.G., Wiedenheft, B., and Doudna, J.A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol Cell *46*, 606-615.
303. Savva, R., and Pearl, L.H. (1995). Cloning and expression of the uracil-DNA glycosylase inhibitor (UGI) from bacteriophage PBS-1 and crystallization of a uracil-DNA glycosylase-UGI complex. Proteins *22*, 287-289.
304. Schormann, N., Ricciardi, R., and Chattopadhyay, D. (2014). Uracil-DNA glycosylases-structural and functional perspectives on an essential family of DNA repair enzymes. Protein Sci *23*, 1667-1685.

305. Sczepanski, J.T., Jacobs, A.C., and Greenberg, M.M. (2008). Self-promoted DNA interstrand cross-link formation by an abasic site. J Am Chem Soc *130*, 9646-9647.
306. Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci USA *108*, 10098-10103.
307. Seol, J.H., Shim, E.Y., and Lee, S.E. (2018). Microhomology-mediated end joining: Good, bad and ugly. Mutat Res *809*, 81-87.
308. Serrano-Heras, G., Bravo, A., and Salas, M. (2008). Phage phi29 protein p56 prevents viral DNA replication impairment caused by uracil excision activity of uracil-DNA glycosylase. Proc Natl Acad Sci U S A *105*, 19044-19049.
309. Serrano-Heras, G., Ruiz-Maso, J.A., del Solar, G., Espinosa, M., Bravo, A., and Salas, M. (2007). Protein p56 from the Bacillus subtilis phage phi29 inhibits DNA-binding ability of uracil-DNA glycosylase. Nucleic Acids Res *35*, 5393-5401.
310. Severinov, K., Kashlev, M., Severinova, E., Bass, I., McWilliams, K., Kutter, E., Nikiforov, V., Snyder, L., and Goldfarb, A. (1994). A non-essential domain of Escherichia coli RNA polymerase required for the action of the termination factor Alc. J Biol Chem *269*, 14254-14259.
311. Shin, J., Jiang, F., Liu, J.J., Bray, N.L., Rauch, B.J., Baik, S.H., Nogales, E., Bondy-Denomy, J., Corn, J.E., and Doudna, J.A. (2017). Disabling Cas9 by an anti-CRISPR DNA mimic. Sci Adv *3*, e1701620.
312. Shuman, S., and Glickman, M.S. (2007). Bacterial DNA repair by non-homologous end joining. Nat Rev Microbiol *5*, 852-861.
313. Signer, E.R., and Weil, J. (1968). Recombination in bacteriophage lambda. I. Mutants deficient in general recombination. J Mol Biol *34*, 261-271.
314. Simmon, V.F., and Lederberg, S. (1972). Degradation of bacteriophage lambda deoxyribonucleic acid after restriction by Escherichia coli K-12. J Bacteriol *112*, 161-169.
315. Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. Genome Res *19*, 1117-1123.
316. Singleton, M.R., Dillingham, M.S., Gaudier, M., Kowalczykowski, S.C., and Wigley, D.B. (2004). Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. Nature *432*, 187-193.
317. Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J *30*, 1335-1342.
318. Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-

mediated CRISPR immunity in *Streptococcus thermophilus*. EMBO J *32*, 385-394.
319. Slavik, K.M., Morehouse, B.R., Ragucci, A.E., Zhou, W., Ai, X., Chen, Y., Li, L., Wei, Z., Bahre, H., Konig, M.*, et al.* (2021). cGAS-like receptors sense RNA and control 3'2'-cGAMP signalling in Drosophila. Nature *597*, 109-113.
320. Slupphaug, G., Mol, C.D., Kavli, B., Arvai, A.S., Krokan, H.E., and Tainer, J.A. (1996). A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. Nature *384*, 87-92.
321. Smith, G. (1983). General recombination In: Hendrix RW, Roberts JW, Stahl FW, Weisberg RA (eds) Lambda II (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).
322. Smith, G.R. (2012). How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. Microbiol Mol Biol Rev *76*, 217-228.
323. Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., Arrieta, J.M., and Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci U S A *103*, 12115-12120.
324. Sommer, N., Depping, R., Piotrowski, M., and Ruger, W. (2004). Bacteriophage T4 alpha-glucosyltransferase: a novel interaction with gp45 and aspects of the catalytic mechanism. Biochem Biophys Res Commun *323*, 809-815.
325. Spies, M., Bianco, P.R., Dillingham, M.S., Handa, N., Baskin, R.J., and Kowalczykowski, S.C. (2003). A molecular throttle: the recombination hotspot chi controls DNA translocation by the RecBCD helicase. Cell *114*, 647-654.
326. Spies, M., Dillingham, M.S., and Kowalczykowski, S.C. (2005). Translocation by the RecB motor is an absolute requirement for chi-recognition and RecA protein loading by RecBCD enzyme. J Biol Chem *280*, 37078-37087.
327. Sriprakash, K.S., Lundh, N., Huh, M.-O., and Radding, C.M. (1975). The specificity of lambda exonuclease. Interactions with single-stranded DNA. J Biol Chem *250*, 5438-5445.
328. Stanley, S.Y., Borges, A.L., Chen, K.H., Swaney, D.L., Krogan, N.J., Bondy-Denomy, J., and Davidson, A.R. (2019). Anti-CRISPR-Associated Proteins Are Crucial Repressors of Anti-CRISPR Transcription. Cell *178*, 1452-1464 e1413.
329. Stanley, S.Y., and Maxwell, K.L. (2018). Phage-Encoded Anti-CRISPR Defenses. Annu Rev Genet *52*, 445-464.
330. Sternberg, S.H., LaFrance, B., Kaplan, M., and Doudna, J.A. (2015). Conformational control of DNA target cleavage by CRISPR-Cas9. Nature *527*, 110-113.
331. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62-67.
332. Stewart, F.J., Panne, D., Bickle, T.A., and Raleigh, E.A. (2000). Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. J Mol Biol *298*, 611-622.

333. Stivers, J.T. (2004). Site-specific DNA damage recognition by enzyme-induced base flipping. Prog Nucleic Acid Res Mol Biol *77*, 37-65.
334. Stokar-Avihail, A., Fedorenko, T., Hor, J., Garb, J., Leavitt, A., Millman, A., Shulman, G., Wojtania, N., Melamed, S., Amitai, G.*, et al.* (2023). Discovery of phage determinants that confer sensitivity to bacterial immune systems. Cell *186*, 1863-1876 e1816.
335. Strotskaya, A., Savitskaya, E., Metlitskaya, A., Morozova, N., Datsenko, K.A., Semenova, E., and Severinov, K. (2017). The action of Escherichia coli CRISPR-Cas system on lytic bacteriophages with different lifestyles and development strategies. Nucleic Acids Res *45*, 1946-1957.
336. Sukhanova, M.V., Khodyreva, S.N., Lebedeva, N.A., Prasad, R., Wilson, S.H., and Lavrik, O.I. (2005). Human base excision repair enzymes apurinic/apyrimidinic endonuclease1 (APE1), DNA polymerase beta and poly(ADP-ribose) polymerase 1: interplay between strand-displacement DNA synthesis and proofreading exonuclease activity. Nucleic Acids Res *33*, 1222-1229.
337. Sun, B., Latham, K.A., Dodson, M.L., and Lloyd, R.S. (1995). Studies on the catalytic mechanism of five DNA glycosylases. Probing for enzyme-DNA imino intermediates. J Biol Chem *270*, 19501-19508.
338. Sun, J.Z., Julin, D.A., and Hu, J.S. (2006). The nuclease domain of the Escherichia coli RecBCD enzyme catalyzes degradation of linear and circular single-stranded and double-stranded DNA. Biochemistry *45*, 131-140.
339. Sun, L., Wu, J., Du, F., Chen, X., and Chen, Z.J. (2013). Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. Science *339*, 786-791.
340. Sung, P. (1994). Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. Science *265*, 1241-1243.
341. Suttle, C.A. (2007). Marine viruses--major players in the global ecosystem. Nat Rev Microbiol *5*, 801-812.
342. Swinton, D., Hattman, S., Benzinger, R., Buchanan-Wollaston, V., and Beringer, J. (1985). Replacement of the deoxycytidine residues in Rhizobium bacteriophage RL38JI DNA. FEBS Lett *184*, 294-298.
343. Takahashi, I., and Marmur, J. (1963). Replacement of thymidylic acid by deoxyuridylic acid in the deoxyribonucleic acid of a transducing phage for Bacillus subtilis. Nature *197*, 794-795.
344. Tal, N., Millman, A., Stokar-Avihail, A., Fedorenko, T., Leavitt, A., Melamed, S., Yirmiya, E., Avraham, C., Brandis, A., Mehlman, T.*, et al.* (2022). Bacteria deplete deoxynucleotides to defend against bacteriophage infection. Nat Microbiol *7*, 1200-1209.
345. Tal, N., Morehouse, B.R., Millman, A., Stokar-Avihail, A., Avraham, C., Fedorenko, T., Yirmiya, E., Herbst, E., Brandis, A., Mehlman, T.*, et al.* (2021). Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. Cell *184*, 5728-5739 e5716.

346. Talpaert-Borle, M. (1987). Formation, detection and repair of AP sites. Mutat Res *181*, 45-56.
347. Tang, T.H., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., Bachellerie, J.P., and Huttenhofer, A. (2005). Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. Mol Microbiol *55*, 469-481.
348. Tao, P., Wu, X., and Rao, V. (2018). Unexpected evolutionary benefit to phages imparted by bacterial CRISPR-Cas9. Sci Adv *4*, eaar4134.
349. Taylor, A.F., and Smith, G.R. (1985). Substrate specificity of the DNA unwinding activity of the RecBC enzyme of *Escherichia coli*. J Mol Biol *185*, 431-443.
350. Taylor, A.F., and Smith, G.R. (1992). RecBCD enzyme is altered upon cutting DNA at a chi recombination hotspot. Proc Natl Acad Sci U S A *89*, 5226-5230.
351. Taylor, A.F., and Smith, G.R. (2003). RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. Nature *423*, 889-893.
352. Tesson, F., Herve, A., Mordret, E., Touchon, M., d'Humieres, C., Cury, J., and Bernheim, A. (2022). Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun *13*, 2561.
353. Thavalingam, A., Cheng, Z., Garcia, B., Huang, X., Shah, M., Sun, W., Wang, M., Harrington, L., Hwang, S., Hidalgo-Reyes, Y.*, et al.* (2019). Inhibition of CRISPR-Cas9 ribonucleoprotein complex assembly by anti-CRISPR AcrIIC2. Nat Commun *10*, 2806.
354. Thomas, J.A., Orwenyo, J., Wang, L.X., and Black, L.W. (2018). The Odd "RB" Phage-Identification of Arabinosylation as a New Epigenetic Modification of DNA in T4-Like Phage RB69. Viruses *10*.
355. Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., Prill, R.J., Tripathi, A., Gibbons, S.M., Ackermann, G.*, et al.* (2017). A communal catalogue reveals Earth's multiscale microbial diversity. Nature *551*, 457-463.
356. Thompson, P.S., and Cortez, D. (2020). New insights into abasic site repair and tolerance. DNA Repair (Amst) *90*, 102866.
357. Tiemann, B., Depping, R., Gineikiene, E., Kaliniene, L., Nivinskas, R., and Ruger, W. (2004). ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. J Bacteriol *186*, 7262-7272.
358. Tock, M.R., and Dryden, D.T. (2005). The biology of restriction and anti-restriction. Curr Opin Microbiol *8*, 466-472.
359. Touchon, M., Charpentier, S., Clermont, O., Rocha, E.P., Denamur, E., and Branger, C. (2011). CRISPR distribution within the Escherichia coli species is not suggestive of immunity-associated diversifying selection. J Bacteriol *193*, 2460-2467.
360. Touzain, F., Petit, M.A., Schbath, S., and El Karoui, M. (2011). DNA motifs that sculpt the bacterial chromosome. Nat Rev Microbiol *9*, 15-26.

361. Trifinopoulos, J., Nguyen, L.T., von Haeseler, A., and Minh, B.Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res *44*, W232-235.
362. Twort, F.W. (1915). An investigation on the nature of ultra-microscopic viruses. The Lancet *186*, 1241-1243.
363. Uribe, R.V., van der Helm, E., Misiakou, M.A., Lee, S.W., Kol, S., and Sommer, M.O.A. (2019). Discovery and Characterization of Cas9 Inhibitors Disseminated across Seven Bacterial Phyla. Cell Host Microbe *25*, 233-241 e235.
364. van Houte, S., Ekroth, A.K., Broniewski, J.M., Chabas, H., Ashby, B., Bondy-Denomy, J., Gandon, S., Boots, M., Paterson, S., Buckling, A.*, et al.* (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. Nature *532*, 385-388.
365. VanderWal, A.R., Park, J.U., Polevoda, B., Nicosia, J.K., Molina Vargas, A.M., Kellogg, E.H., and O'Connell, M.R. (2023). Csx28 is a membrane pore that enhances CRISPR-Cas13b-dependent antiphage defense. Science *380*, 410-415.
366. Varble, A., Campisi, E., Euler, C.W., Maguin, P., Kozlova, A., Fyodorova, J., Rostol, J.T., Fischetti, V.A., and Marraffini, L.A. (2021). Prophage integration into CRISPR loci enables evasion of antiviral immunity in Streptococcus pyogenes. Nat Microbiol *6*, 1516-1525.
367. Vassallo, C.N., Doering, C.R., Littlehale, M.L., Teodoro, G.I.C., and Laub, M.T. (2022). A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome. Nat Microbiol *7*, 1568-1579.
368. Vlot, M., Houkes, J., Lochs, S.J.A., Swarts, D.C., Zheng, P., Kunne, T., Mohanraju, P., Anders, C., Jinek, M., van der Oost, J.*, et al.* (2018). Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR-Cas effector complexes. Nucleic Acids Res *46*, 873-885.
369. Waite-Rees, P.A., Keating, C.J., Moran, L.S., Slatko, B.E., Hornstra, L.J., and Benner, J.S. (1991). Characterization and expression of the Escherichia coli Mrr restriction system. J Bacteriol *173*, 5207-5219.
370. Walkinshaw, M.D., Taylor, P., Sturrock, S.S., Atanasiu, C., Berge, T., Henderson, R.M., Edwardson, J.M., and Dryden, D.T. (2002). Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. Mol Cell *9*, 187-194.
371. Wang, J., Chen, R., and Julin, D.A. (2000). A single nuclease active site of the Escherichia coli RecBCD enzyme catalyzes single-stranded DNA degradation in both directions. J Biol Chem *275*, 507-513.
372. Wang, L., Chen, S., Xu, T., Taghizadeh, K., Wishnok, J.S., Zhou, X., You, D., Deng, Z., and Dedon, P.C. (2007). Phosphorothioation of DNA in bacteria by dnd genes. Nat Chem Biol *3*, 709-710.
373. Wang, N., Bao, H., Chen, L., Liu, Y., Li, Y., Wu, B., and Huang, H. (2019). Molecular basis of abasic site sensing in single-stranded DNA by the SRAP domain of E. coli yedK. Nucleic Acids Res *47*, 10388-10399.

374. Wang, S., Wan, M., Huang, R., Zhang, Y., Xie, Y., Wei, Y., Ahmad, M., Wu, D., Hong, Y., Deng, Z.*, et al.* (2021). SspABCD-SspFGH Constitutes a New Type of DNA Phosphorothioate-Based Bacterial Defense System. mBio *12*.
375. Wang, X., Kim, Y., Ma, Q., Hong, S.H., Pokusaeva, K., Sturino, J.M., and Wood, T.K. (2010). Cryptic prophages help bacteria cope with adverse environments. Nat Commun *1*, 147.
376. Wang, X., Yao, D., Xu, J.G., Li, A.R., Xu, J., Fu, P., Zhou, Y., and Zhu, Y. (2016). Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. Nat Struct Mol Biol *23*, 868-870.
377. Wang, Z., and Mosbaugh, D.W. (1989). Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. J Biol Chem *264*, 1163-1171.
378. Warren, R.A. (1980). Modified bases in bacteriophage DNAs. Annu Rev Microbiol *34*, 137-158.
379. Watters, K.E., Fellmann, C., Bai, H.B., Ren, S.M., and Doudna, J.A. (2018). Systematic discovery of natural CRISPR-Cas12a inhibitors. Science *362*, 236-239.
380. Wegrzyn, G., Licznerska, K., and Wegrzyn, A. (2012). Phage lambda--new insights into regulatory circuits. Adv Virus Res *82*, 155-178.
381. Weigele, P., and Raleigh, E.A. (2016). Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. Chem Rev *116*, 12655-12687.
382. Weller, G.R., Kysela, B., Roy, R., Tonkin, L.M., Scanlan, E., Della, M., Devine, S.K., Day, J.P., Wilkinson, A., d'Adda di Fagagna, F.*, et al.* (2002). Identification of a DNA nonhomologous end-joining complex in bacteria. Science *297*, 1686-1689.
383. West, S.C. (1994). The processing of recombination intermediates: mechanistic insights from studies of bacterial proteins. Cell *76*, 9-15.
384. Westra, E.R., Pul, U., Heidrich, N., Jore, M.M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L.*, et al.* (2010). H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. Mol Microbiol *77*, 1380-1393.
385. Westra, E.R., van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K.*, et al.* (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol Cell *46*, 595-605.
386. Wiberg, J.S. (1967). Amber mutants of bacteriophage T4 defective in deoxycytidine diphosphatase and deoxycytidine triphosphatase. On the role of 5-hydroxymethylcytosine in bacteriophage deoxyribonucleic acid. J Biol Chem *242*, 5824-5829.
387. Wibley, J.E., Waters, T.R., Haushalter, K., Verdine, G.L., and Pearl, L.H. (2003). Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. Mol Cell *11*, 1647-1659.

388. Wiedenheft, B., van Duijn, E., Bultema, J., Waghmare, S., Zhou, K., Barendregt, A., Westphal, W., Heck, A., Boekema, E., Dickman, M.*, et al.* (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proc Natl Acad Sci USA *108*, 10092-10097.
389. Wigley, D.B. (2013). Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. Nat Rev Microbiol *11*, 9-13.
390. Wilde, J.A., Bolton, P.H., Mazumder, A., Manoharan, M., and Gerlt, J.A. (1989). Characterization of the equilibrating forms of the aldehydic abasic site in duplex DNA by oxygen-17 NMR. J Am Chem Soc *111*, 1894-1896.
391. Wilhelm, K., and Ruger, W. (1992). Deoxyuridylate-hydroxymethylase of bacteriophage SPO1. Virology *189*, 640-646.
392. Wilkinson, M., Troman, L., Wan Nur Ismah, W.A., Chaban, Y., Avison, M.B., Dillingham, M.S., and Wigley, D.B. (2016). Structural basis for the inhibition of RecBCD by Gam and its synergistic antibacterial effect with quinolones. Elife *5*.
393. Williams, J.G., and Radding, C.M. (1981). Partial purification and properties of an exonuclease inhibitor induced by bacteriophage Mu-1. J Virol *39*, 548-558.
394. Williams, M.C., Reker, A.E., Margolis, S.R., Liao, J., Wiedmann, M., Rojas, E.R., and Meeske, A.J. (2023). Restriction endonuclease cleavage of phage DNA enables resuscitation from Cas13-induced bacterial dormancy. Nat Microbiol *8*, 400-409.
395. Wozniak, K.J., and Simmons, L.A. (2022). Bacterial DNA excision repair pathways. Nat Rev Microbiol *20*, 465-477.
396. Wright, W.D., Shah, S.S., and Heyer, W.D. (2018). Homologous recombination and the repair of DNA double-strand breaks. J Biol Chem *293*, 10524-10535.
397. Wu, X., Zhu, J., Tao, P., and Rao, V.B. (2021). Bacteriophage T4 Escapes CRISPR Attack by Minihomology Recombination and Repair. mBio, e0136121.
398. Wyatt, G.R., and Cohen, S.S. (1953). The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. Biochem J *55*, 774-782.
399. Xiong, L., Liu, S., Chen, S., Xiao, Y., Zhu, B., Gao, Y., Zhang, Y., Chen, B., Luo, J., Deng, Z.*, et al.* (2019). A new type of DNA phosphorothioation-based antiviral system in archaea. Nat Commun *10*, 1688.
400. Xiong, X., Wu, G., Wei, Y., Liu, L., Zhang, Y., Su, R., Jiang, X., Li, M., Gao, H., Tian, X.*, et al.* (2020). SspABCD-SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities. Nat Microbiol *5*, 917-928.
401. Xu, T., Yao, F., Zhou, X., Deng, Z., and You, D. (2010). A novel host-specific restriction system associated with DNA backbone S-modification in Salmonella. Nucleic Acids Res *38*, 7133-7141.
402. Yang, H., and Patel, D.J. (2017). Inhibition Mechanism of an Anti-CRISPR Suppressor AcrIIA4 Targeting SpyCas9. Mol Cell *67*, 117-127 e115.

403. Yasmin, T., Azeroglu, B., Cockram, C.A., and Leach, D.R.F. (2021). Distribution of Holliday junctions and repair forks during Escherichia coli DNA double-strand break repair. PLoS Genet *17*, e1009717.
404. Zaremba, M., Dakineviciene, D., Golovinas, E., Zagorskaite, E., Stankunas, E., Lopatina, A., Sorek, R., Manakova, E., Ruksenaite, A., Silanskas, A.*, et al.* (2022). Short prokaryotic Argonautes provide defence against incoming mobile genetic elements through NAD(+) depletion. Nat Microbiol *7*, 1857-1869.
405. Zeng, Z., Chen, Y., Pinilla-Redondo, R., Shah, S.A., Zhao, F., Wang, C., Hu, Z., Wu, C., Zhang, C., Whitaker, R.J.*, et al.* (2022). A short prokaryotic Argonaute activates membrane effector to confer antiviral defense. Cell Host Microbe *30*, 930-943 e936.
406. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A.*, et al.* (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759-771.
407. Zhang, T., Tamman, H., Coppieters 't Wallant, K., Kurata, T., LeRoux, M., Srikant, S., Brodiazhenko, T., Cepauskas, A., Talavera, A., Martens, C.*, et al.* (2022). Direct activation of a bacterial innate immune system by a viral capsid protein. Nature *612*, 132-140.
408. Zhang, Y., Matsuzaka, T., Yano, H., Furuta, Y., Nakano, T., Ishikawa, K., Fukuyo, M., Takahashi, N., Suzuki, Y., Sugano, S.*, et al.* (2017). Restriction glycosylases: involvement of endonuclease activities in the restriction process. Nucleic Acids Res *45*, 1392-1403.
409. Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., and Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. Nature *515*, 147-150.
410. Zhu, J.K. (2009). Active DNA demethylation mediated by DNA glycosylases. Annu Rev Genet *43*, 143-166.
411. Zhu, Y., Gao, A., Zhan, Q., Wang, Y., Feng, H., Liu, S., Gao, G., Serganov, A., and Gao, P. (2019). Diverse Mechanisms of CRISPR-Cas9 Inhibition by Type IIC Anti-CRISPR Proteins. Mol Cell.
412. Zhu, Y., Yin, S., and Li, Z. (2023). Mechanism of inhibition of CRISPR-Cas9 by anti-CRISPR protein AcrIIC1. Biochem Biophys Res Commun *654*, 34-39.
413. Zimmermann, L., Stephens, A., Nam, S.Z., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, J., Lupas, A.N., and Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol *430*, 2237-2243.