Costs and benefits of diversity-generating immune mechanisms

Submitted by David Sünderhauf to the University of Exeter as a thesis for the degree of Masters by Research in Biological Sciences, September 2017
Version with Minor Corrections resubmitted March 2018

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any University.

Signature: ……

Abstract

Organisms across the tree of life have evolved diversity-generating immune mechanisms (DGMs) to counteract selective pressures imposed by their parasites. Increased host diversity has a major impact on parasite epidemics as well as host evolution. Being virtually ubiquitous, bacteria and their predators, bacteriophage (phage), are essential to every ecological niche on earth and key players in industrial and healthcare applications. Bacterial DGMs include CRISPR-Cas and Restriction-Modification (RM) shufflons. Type I RM methylates self-DNA and cleaves unmethylated invasive DNA, however phage can escape from this response by becoming methylated themselves. Shufflons recombine genes coding for the RM specificity subunit, creating population-level diversity in recognition sequences; this is thought to limit phage escape. We investigate the *Mycoplasma pulmonis* Mpu shufflon, which has the capacity to generate 30 different specificity subunits, of which we predict at least 12 to be functional. We create a model system by adapting the Mpu shufflon for expression in *Pseudomonas aeruginosa* PA14. Transforming a CRISPR-deficient PA14 host with RM, we uncover large autoimmune costs when inducing a novel RM system with only limited benefits of low-level phage resistance. When expressed together, CRISPR-Cas and RM provide PA14 with complete resistance against most *Pseudomonas* phages tested and partial resistance against lipopolysaccharide-specific phage LMA2. Surprisingly, the RM restriction subunit is not an essential component for this effect; the mechanistic basis of this synergistic interaction between DNA methylation and CRISPR-Cas systems requires further investigation. The lack of detectable spacer acquisition, required for CRISPR-Cas to effectively target the infecting phage, suggests these effects are likely due to altered host gene expression that in turn impacts the ability of phage to infect. Future studies need to address questions about the molecular basis of resistance in this model system.

Acknowledgments

Firstly, I would like to thank my supervisor Dr Edze Westra who not only supported me throughout my Master's project, but also encouraged me to pursue an MbyRes degree in his laboratory in the first place. I could not have wished for a more proactive supervisor who continued to challenge, inspire, and set time aside to help me from start to finish. Secondly, it is thanks to Dr Mariann Landsberger that I don't feel like a beginner in the laboratory anymore, and can successfully pretend to know my way around to any new students. She also supported me throughout my whole degree, taking great interest in a project that wasn't hers whilst giving very helpful academic, technical, and personal advice.

Furthermore, I would like to thank other members of the Westra Group for their general comradeship and making the group a joy to work in. Thanks to Dr Stineke van Houte (who rescued me from going crazy over *in silico* work by leading me through a lab side-project, and apparently wasn't put off enough to stop her from wanting to be my PhD supervisor), Jack (who was always happy to help and chat, and whom I still want to help with his evolution experiment), Dan (my fellow MRes student), the two Ellies, Jenny, Sean, Clare (who kindly read through my thesis draft), Devi, and Hélène. Furthermore, thanks to all Buckling group lab members (including Prof Angus Buckling, whom I had to surprise with the information that he was my second supervisor, and who gave helpful insights throughout), other ESI lab users and anyone I may have forgotten. I am sad to be abandoning Restriction-Modification and this project, but grateful that I will be able to spend another few years in this laboratory with these great people.

Lastly, thanks to my friends and especially my fiancée Sophie for reminding me what really matters and keeping me sane throughout, and who made living in Cornwall so much better.

Table of Contents

List of Tables and Figures

Author's Declaration

Chapter I is a collaborative effort that has since been published as a literature review. I wrote the initial review proposal and researched data and papers for inclusion in the article. I wrote first drafts of the Introduction, Targeted DGMs, Cost of DGMs, and DGM-associated Coevolution sections, and created figure 1, and parts of figures 2 and 3. Furthermore, I was involved in content and figure discussion throughout the writing process. The article can be found as follows:

Westra, E. R., Sünderhauf, D., Landsberger, M., & Buckling, A. (2017). Mechanisms and consequences of diversity-generating immune strategies. *Nature Reviews Immunology*. https://doi.org/10.1038/nri.2017.78

I would like to thank and acknowledge my co-authors for their work and do not claim sole ownership over this article and thesis chapter.

Abbreviations

Acr	anti-CRISPR
AID	activation-induced cytidine deaminase
ARD	arms-race dynamics
ATP	adenosine triphosphate
BCR	B cell receptor
Cas	CRISPR-associated
CFU	colony-forming units
CRISPR	clustered regularly interspaced short palindromic repeats
CSR	class-switch recombination
DGM	diversity-generating mechanism
DNA	deoxyribonucleic acid
EOP	efficiency of plaquing
EOT	efficiency of transformation
FSD	fluctuating selection dynamics
gDNA	genomic DNA
GM50	50 µg/ml gentamycin
Hsd-	host specificity for DNA -
-S	-specificity
-R	-restriction
-M	-methylation/modification
hrs	*hsd* recombination site
kb	kilobases (1000 base pairs)
LPS	lipopolysaccharide
LRR	leucine-rich repeat
MCS	multiple cloning site
MOI	multiplicity of infection
MHC	major histocompatibility complex
nt	nucleotide
PCR	polymerase chain reaction
pfu	plaque-forming units
phage	bacteriophage
qPCR	quantitative PCR
RA	restriction alleviation

RAG	recombination-activating gene
RBS	ribosome binding site
RISC	RNA-induced silencing complex
RM	Restriction-Modification
RNA	ribonucleic acid
RNAi	RNA interference
SNP	single-nucleotide polymorphism
TCR	T cell receptor
TRD	target recognition domain
vip	vipareetus
viRNA	virus-derived small interfering RNA
VLR	variable lymphocyte receptor
VSG	variant-specific glycoprotein

General Introduction

Where unlimited resources and space exist, plants, animals, and other organisms can theoretically thrive indefinitely. However, most living beings are subject to predation or parasitism. In all domains of life, interactions of hosts with their parasites are crucial to evolutionary dynamics and can define ecosystem functions. In particular, parasites are known to sweep through populations with very low diversity, such as crop monocultures (Zhu *et al.*, 2000).

Bacteria and their parasites bacteriophage (phage) comprise an excellent model system to study host-pathogen interactions thanks to their relative simplicity and the ease with which these dynamics can be influenced in a very controlled way in laboratory environments. Additionally, they are intrinsically interesting owing to them being the most abundant organisms on the planet with key ecological functions such as nutrient cycling. Understanding the coevolutionary dynamics of bacteria and phage can lead to applications in a range of fields including agriculture, the food industry, and healthcare. As the most abundant life-form on this planet (Breitbart and Rohwer, 2005), phage have the capacity to limit bacterial growth in virtually any environment. Similar to our own viruses, phage inject their DNA into a bacterial host to either remain latent over a period of time or to hijack the host's replication machinery and generate progeny phage. Generally, bacteria can evolve resistance to phage by preventing phage adsorption, through mutation, masking or downregulating the phage receptor, or by one of several more sophisticated immune responses (Labrie, Samson and Moineau, 2010). The most prevalent immune mechanism, Restriction-Modification (RM), functions by methylating bacterial self-DNA, and cleaving unmethylated invasive DNA (Fig i). Alternatively, CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated) can integrate short phage DNA sequences into its own genome as spacers, and later use their transcripts as guides to specifically cleave target DNA with the same sequence. As innate and adaptive immune responses respectively, RM and CRISPR-Cas are associated with constitutive or inducible costs respectively and distinct evolutionary dynamics. This means that immunity through either response is driven by phage exposure and nutrient availability (Westra *et al.*, 2015).

As phage can escape from RM with relative ease when they become methylated by chance, RM is predicted to provide the largest benefit to bacteria when they

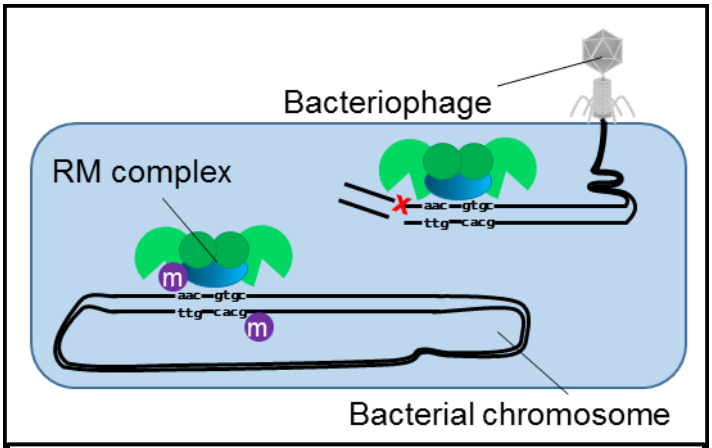


Figure i. Self/non-self discrimination through Restriction-Modification.
A simplified diagram of a Type I RM system selectively cleaving foreign DNA. Bacterial DNA is methylated (m) at recognition sites by the RM complex, while foreign bacteriophage DNA is cleaved (red x) when the RM complex recognises unmethylated sites.

are invading a new ecosystem, whereas phage in an environment with established bacterial populations are more likely to be methylated and therefore invisible to this immune response (Korona and Levin, 1993). Bearing this in mind, RM systems nonetheless provide a long-term advantage to their hosts over strains without this immune system (Sneppen *et al.*, 2015). Benefits to bacteria can be further mitigated by autoimmunity, which is toxic and occurs when RM accidentally cleaves the bacterial genome (Pleška *et al.*, 2016). Furthermore, RM can also have roles beyond immunity that are not directly attributed to their DNA cleaving function (reviewed in Vasu and Nagaraja, 2013), which can complicate predictions of costs or benefits. These include: directing the rate of evolution through modulating genetic variation; facilitating recombination; regulating gene expression through methylation patterns; and in some cases even metabolic functions of phage. Additionally, RM can trigger altruistic apoptosis (Nagamalleswari *et al.*, 2017). Although not as common, CRISPR-Cas can fulfil alternative roles, too (reviewed in Westra, Buckling and Fineran, 2014). These include gene regulation as well as bacterial virulence. Additionally, both immune systems act as selfish genetic elements in some cases. Data has shown that RM and CRISPR-Cas are not only compatible immune responses (Dupuis *et al.*, 2013), but that RM might actively aid CRISPR spacer acquisition (Hynes, Villion and Moineau, 2014).

Our model organism *Pseudomonas aeruginosa* is a gram-negative bacterium found in soil, water, and animals, which is equipped with a Type I-F CRISPR-Cas system. In humans, *P. aeruginosa* is an opportunistic pathogen of the lung and other tissues (Lyczak, Cannon and Pier, 2000), and in soil it can cause plant

disease. Its interactions with phage (especially DMS3vir, a lytic *Pseudomonas* phage) have been thoroughly studied (Zegans *et al.*, 2009; Chabas *et al.*, 2016; van Houte *et al.*, 2016; Morley *et al.*, 2017). In high nutrient conditions and exposed to a large amount of phage, *P. aeruginosa* evolves phage resistance by surface modification, through mutational loss of the pilus. When nutrients are scarce or the phage load is low, CRISPR-Cas is the preferred mechanism of resistance (Westra *et al.*, 2015). *P. aeruginosa* PA14, the strain used for our experiments, does not have a native RM system.

In this thesis, I discuss the benefits of RM shufflon-associated diversity as well as costs and benefits of RM expression in bacteria. First, I review diversity-generating immune mechanisms across the entire tree of life and discuss what diversity-dependent predictions on host/parasite coevolution can be made (Chapter I). I then focus on Type I RM and its unique capacity for evolution of diversity, examine the *Mycoplasma pulmonis* Mpu shufflon for its recombinatory properties, and predict which diversity subunits it can generate (Chapter II). I adapt this shufflon for expression in *Pseudomonas aeruginosa* PA14 to establish an experimental system for investigations of RM diversity as well as coordinated impact of CRISPR-Cas and RM on bacteria-phage coevolution. I explore RM expression in a CRISPR-deficient host and pose questions about the costs and benefits associated with inducing novel RM systems (Chapter III). Finally, I express RM in conjunction with CRISPR-Cas in an attempt to ascertain the joint impact of these bacterial immune systems on phage resistance (Chapter IV).

Chapter I: Mechanisms and consequences of diversity-generating immune strategies

Edze R. Westra, David Sünderhauf, Mariann Landsberger, and Angus Buckling
Nature Reviews Immunology 2017. doi:10.1038/nri.2017.78

Abstract

Species from all five kingdoms of life have evolved sophisticated mechanisms to generate diversity in genes that are involved in host–pathogen interactions, conferring reduced levels of parasitism to both individuals and populations. Here, we highlight unifying concepts that underpin these evolutionarily unrelated diversity-generating mechanisms (DGMs). We discuss the mechanisms of and selective forces acting on these diversity-generating immune strategies, as well as their epidemiological and co-evolutionary consequences. We propose that DGMs can be broadly classified into two classes — targeted and untargeted DGMs — which generate different levels of diversity with important consequences for host–parasite co-evolution.

Introduction

Over the past few decades, theoretical, correlational and experimental studies have linked increased host diversity to lower levels of pathogen infection (reviewed in Keesing, Holt and Ostfeld, 2006). This observation initially came from agriculture, where it is known as the monoculture effect (Elton, 1958), but also applies to animals. For example, natural populations with low levels of genetic diversity (such as endangered species (O'Brien *et al.*, 1985) and inbred animals (Acevedo-Whitehouse *et al.*, 2003)) tend to have higher pathogen loads, and experimental increases in host diversity result in decreased pathogen abundance (Altermatt and Ebert, 2008; van Houte *et al.*, 2016).

Consistent with these observations, many organisms across the five kingdoms of life have evolved sophisticated diversity-generating mechanisms (DGMs) that provide benefits to both individuals and populations in the presence of infectious disease. In this Opinion article, we first summarize the mechanistic basis of the

most well-known DGMs, which have been studied in detail over the past few decades by immunologists, geneticists and molecular biologists, and highlight unifying concepts across the different immune strategies. We then link these molecular data with insights from evolutionary biology and epidemiology to discuss when and why different DGMs are particularly beneficial and their broader implications for disease epidemics and host–pathogen co-evolution (that is, their reciprocal adaptation). Understanding the link between these molecular and macroscopic processes can guide future molecular immunology research into the scale of the host diversity of resistance alleles in space and time. Key questions to address involve how much diversity can be generated in one individual and between individuals as well as how such diversity is maintained over time. An increased knowledge of resistance-allele diversity could help to predict the emergence and spread of infectious diseases.

We propose that DGMs can be broadly classified into two groups: those that function across the entire host genome (for example, mutation and sexual reproduction; referred to here as untargeted DGMs) and those that are targeted to specific loci involved in host–pathogen interactions (for example, V(D)J recombination and prokaryotic CRISPR–Cas; referred to here as targeted DGMs). Untargeted DGMs tend to generate relatively low levels of diversity that pathogens can overcome by genetic mutation or recombination, whereas the higher levels of diversity that are typically generated by targeted DGMs drive the evolution of more sophisticated anti-DGM strategies as part of an arms race between the host and its pathogens.

Untargeted DGMs

Untargeted DGMs, which include any mechanism that creates diversity across the entire genome of a host species, are widespread. We discuss how germline mutation and sexual reproduction generate untargeted diversity and detail the selective advantage of these mechanisms when hosts are exposed to pathogens.

Germline mutation. The simplest mechanism for generating genetic diversity between parents and their offspring is through germline mutations, such as single-nucleotide polymorphisms (SNPs), indels, gene duplications, inversions and transposable element insertions (Fig. I/1a). Mutation rates vary between species

and between strains of the same species. For example, SNPs are typically generated at frequencies from 10^{-7} to 10^{-10} base substitutions per nucleotide per generation, which is well above the biochemical limit to proofreading during DNA replication. Whereas it has been suggested that variation in mutation rates between species may result from genetic drift (Lynch *et al.*, 2016), the selection of beneficial mutations may also help to explain variation within species (Taddei *et al.*, 1997). Theoretical analyses predict that host–pathogen co-evolution can select for increased mutation rates (Pal *et al.*, 2007; Zaman *et al.*, 2014) that allow for the more rapid generation of resistance mutations. Consistent with this prediction, bacteria co-evolving with bacteriophage (phage) frequently evolved 10- to 100-fold increased mutation rates (mutator strains), and these strains were

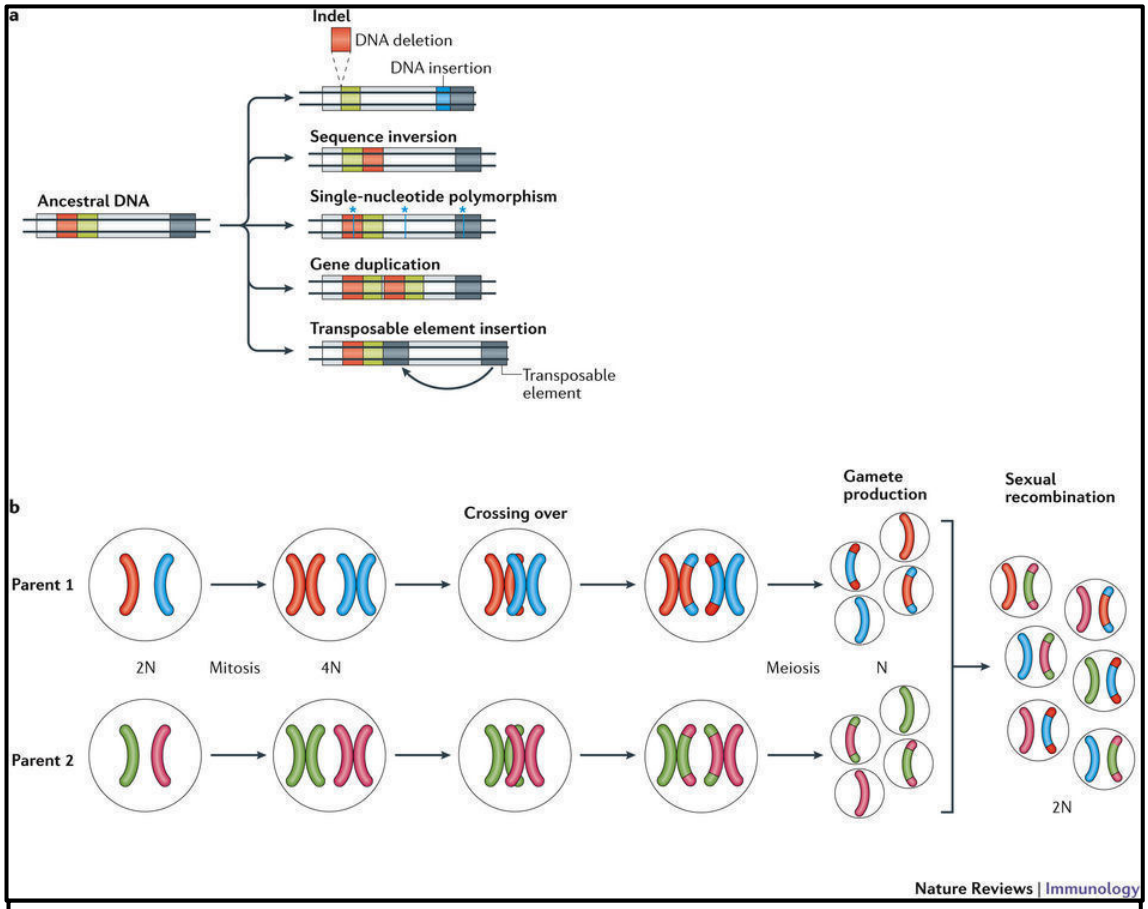


Figure I/1. Untargeted diversity-generating mechanisms.
a Schematic overview of the five main mechanisms of mutation that can increase genetic diversity within a population: DNA insertion or deletion (indel), sequence inversion, single-nucleotide polymorphism, gene duplication and transposable element insertion. **b** Sexual reproduction creates novel allele combinations within a population by crossing over (the exchange of genetic material between homologous chromosomes) during gamete production and by combining novel alleles (in the case of biparental sex).

more likely to drive phage to extinction compared with non-mutator strains (Pal *et al.*, 2007). Selection for mutator strains increases upon exposure to multiple phage (Wielgoss *et al.*, 2016), which may drive the high frequencies of these strains in natural (LeClerc *et al.*, 1996; Matic *et al.*, 1997) and clinical environments (Oliver *et al.*, 2000).

Although such germline mutations are often thought of as random, mutation frequencies tend to be distributed non-uniformly across genomes as a result of nucleotide composition and genomic context (Lynch, 2010; Lee *et al.*, 2012), gene transcription (Herman and Dworkin, 1971; Datta and Jinks-Robertson, 1995) and chromatin organization (Wolfe, Sharp and Li, 1989; Hardison *et al.*, 2003; Makova and Hardison, 2015). The existence of mutational hot and cold spots across genomes has selected for genes that are involved in rapid adaptation (for example, genes associated with pathogen defence) to be located in hot spots, whereas genes that are involved in conserved processes (such as transcription and translation) reside in cold spots (Chuang and Li, 2004).

Sexual reproduction. A more sophisticated form of untargeted DGM is sexual reproduction. Sex generates population-level diversity through recombination during meiosis, which breaks up existing allele combinations and generates novel ones, and (in the case of biparental sexual reproduction) by combining alleles from different individuals in a population during cross-fertilization (Fig. I/1b). In theory, the maintenance of sex as a means of reproduction can be driven by a need to escape pathogens (Hamilton, Axelrod and Tanese, 1990; Lively, 2010a). Specifically, if pathogens are specialized with respect to host genotypes, then sexual reproduction can generate rare genotypes that are susceptible to infection with fewer pathogens in the short term. Many correlational and experimental studies support this idea (reviewed in Lively and Morran, 2014). For example, analyses of the distributions of sexual and parthenogenic forms of a freshwater snail (*Potamopyrgus antipodarum*) show that sexual forms are more common in habitats with high pathogen densities and vice versa (Lively, 1987; King and Lively, 2009; King *et al.*, 2009). In mixed populations, sexual forms tend to be less frequently infected than asexual forms (Vergara, Jokela and Lively, 2014). In *Caenorhabditis elegans*, it has been shown that presence of the bacterial pathogen *Serratia marcescens* selects for outcrossing (Morran *et al.*, 2011). As is the case

for germline mutation, recombination hot spots during sexual reproduction generate relatively large amounts of diversity at genes that are associated with pathogen immunity (for example, MHC genes (Paigen and Petkov, 2010)).

To increase the pathogen resistance of their offspring, females may prefer to mate with males who have optimal resistance alleles (Hamilton and Zuk, 1982). Non-random mating based on MHC genes has been shown for several species (reviewed in Kamiya *et al.*, 2014), including mouse and fish (Potts, Manning and Wakeland, 1991; Reusch *et al.*, 2001). Such mating preferences could be based on chemosensory signals associated with MHC molecules (Leinders-Zufall *et al.*, 2004). Other organisms, such as social insects, are thought to lack this mating preference; instead, females may prefer to mate with multiple males to increase heterogeneity among offspring, presumably to decrease disease spread (Baer and Schmid-Hempel, 1999, 2001, 2003).

Unifying concepts. These examples illustrate how untargeted DGMs generate diversity across genomes and detail their benefits in the presence of infectious disease. However, as these untargeted mechanisms also generate diversity at loci that are not involved in host–pathogen interactions, they can enhance adaptation to other selective pressures, such as environmental change. Therefore, the evolution of untargeted DGMs is not driven by pathogens alone (de Visser and Elena, 2007). For example, germline mutation has an important role in adaptation in general, and increased mutation rates are commonly observed in the absence of pathogens (Sniegowski, Gerrish and Lenski, 1997; Taddei *et al.*, 1997; Giraud *et al.*, 2001). Likewise, sexual reproduction has an important role in adaptation (Morran, Parmenter and Phillips, 2009) by reducing clonal interference (competition between genotypes that carry different advantageous mutations) and breaking the linkage between beneficial and deleterious mutations (McDonald, Rice and Desai, 2016), and it helps to maintain population fitness through sexual selection of the fittest males (Lumley *et al.*, 2015).

Targeted DGMs

Unlike mutation and sexual reproduction, which generate diversity between generations (with the exception of somatic mutations (Forsberg, Gisselsson and

Dumanski, 2017)) and across entire genomes, targeted DGMs can generate diversity upon pathogen exposure specifically in genetic loci that are important for host–pathogen interactions (Fig. I/2). Targeted DGMs generate higher levels of diversity than untargeted DGMs, and this diversity can be generated both between and within hosts.

Vertebrate adaptive immunity. The adaptive immune response of jawed vertebrates involves arguably the most thoroughly studied DGMs (Rast and Litman, 1994). The main sources of diversity are B cells and T cells, which generate large repertoires of B cell receptors (BCRs) and antibodies, and T cell receptors (TCRs), respectively, that can interact with diverse pathogen antigens (Weinstein *et al.*, 2009; Jiang *et al.*, 2011). As outlined below, the adaptive immune response generates both within- and between-host diversity in the resistance alleles encoding these receptors. Note that — aside from the direct transfer of antibodies from mother to offspring — only the propensity to generate diversity is heritable in this case, not the diverse BCR and TCR resistance alleles themselves (reviewed in Boulinier and Staszewski, 2008).

One of the key mechanisms in this diversity-generating process is V(D)J recombination, which acts on multi-copy V (variable), J (joining) and sometimes D (diversity) gene segments by semi-randomly linking together single copies of each to generate unique variable antigen-recognition domains (Early *et al.*, 1980) (Fig. I/2a). This DGM involves many enzymes, including the hairpin-forming transposase recombination-activating gene 1 (RAG1) and its cofactor RAG2 (Schatz, Oettinger and Baltimore, 1989). High-resolution models of the RAG1–RAG2 protein complex (Kim *et al.*, 2015) help to explain why mutations in these genes can cause severe combined immunodeficiency (Schwarz *et al.*, 1996). By contrast, in species such as chicken, almost all diversity comes from gene conversion, in which gene segments in the antibody variable region recombine with gene segments from pseudogenes (Maizels, 1987).

Following V(D)J recombination, the primary antibody repertoire produced by B cells goes through a process known as affinity maturation, which is mediated by activation-induced cytidine deaminase (AID; also known as AICDA). This process takes place in germinal centres, where B cells undergo repeated cycles of mutation (known as somatic hypermutation) in the variable regions of immunoglobulin genes and selection based on affinity for antigen (Allen *et al.*, 2007; Victora *et al.*,

2010; Gitlin *et al.*, 2015). Somatic hypermutation can increase antibody affinity by orders of magnitude and, for example, is essential for generating broadly neutralizing antibodies against influenza virus (Pappas *et al.*, 2014). Although some of the antibody diversity that is generated is inevitably lost during selection as a result of clonal expansion (Gitlin, Shulman and Nussenzweig, 2014), the extent of this loss is variable, and it has been shown that antibodies with different affinities for the same antigen can coexist in the same host (Kuraoka *et al.*, 2016; Tas *et al.*, 2016). In addition to mutation of the variable sequence, antibodies also undergo class-switch recombination (CSR), whereby variable regions are linked to genes encoding alternative constant regions (encoding the non-variable parts of the antibody), which results in a switch of antibody class and therefore interactions with different effectors. The type of switch is determined by cytokine levels as well as genetic factors (Horns *et al.*, 2016).

Lampreys and hagfish are primitive jawless vertebrates that lack BCRs, TCRs and MHC molecules but encode an alternative adaptive immune system that

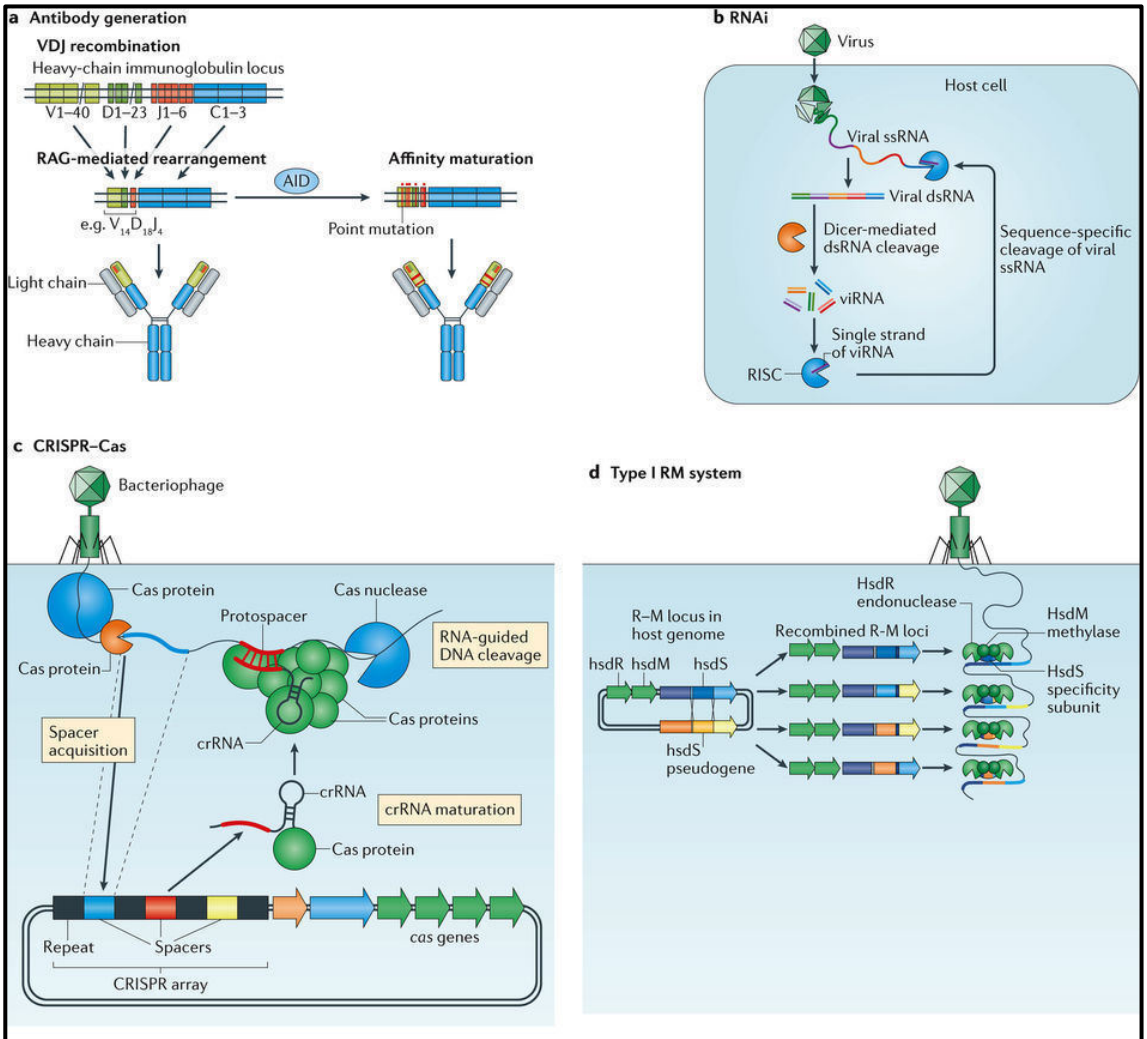


Figure I/2.

Figure I/2. Targeted diversity-generating mechanisms.

a To form an antibody heavy chain gene, one of each of multiple variable (V), diversity (D) and joining (J) segments are joined together and linked to one constant (C) segment, which encodes the constant region. During such V(D)J recombination, V, D and J segments are semi-randomly linked together by the enzymes recombination-activating gene 1 (RAG1) and RAG2, resulting in diverse heavy chain antibody genes. The same process is involved in light chain gene diversification (not shown), but here, only V and J segments recombine. During affinity maturation, activation-induced cytidine deaminase (AID) induces point mutations at cytosine bases within the immunoglobulin variable region, which further increases antibody variation. **b** After a virus invades a cell, Dicer cleaves double-stranded (ds) viral RNA and processes it into virus-derived small interfering RNAs (viRNAs) spanning 21–24 nucleotides in length. A single strand of viRNA gets incorporated into the RNA-induced silencing complex (RISC), which targets and degrades complementary single-stranded (ss) viral RNA. **c** Upon phage infection of a bacterial host, a short phage sequence is incorporated into the bacterial CRISPR array as a spacer. CRISPR transcripts are processed into CRISPR RNA (crRNA), which associates with Cas proteins. The crRNA–Cas ribonucleoprotein complex binds and cleaves complementary nucleic acid sequences (in this example, DNA) in the phage, resulting in immunity. **d** Type I restriction–modification (RM) systems encode a restriction endonuclease (HsdR), a methylase (HsdM) and a specificity subunit (HsdS). HsdM methylates recognition sequences on the bacterial genome, and HsdR cleaves the same sequences when unmodified (for example, in phage DNA). RM shufflons have a second pseudogene copy of hsdS (and in some cases, up to six pseudogene copies), which can recombine with the main *hsdS* gene, resulting in altered sequence specificity of the RM system.

uses antigen receptors termed variable lymphocyte receptors (VLRs) (Pancer *et al.*, 2004). In their mature form, VLRs are composed of up to eight highly variable leucine-rich repeat (LRR) modules flanked by conserved LRR modules and other conserved sequence elements. Three types of VLR have been identified that are expressed by separate lymphocyte lineages (reviewed in Boehm *et al.*, 2012). Lymphocytes somatically diversify their incomplete germline VLR alleles through gene conversion to assemble mature VLRs that contain a unique combination of variable LRR modules derived from the hundreds of different LRR modules that

flank the germline VLR allele. The diversity, specificity and affinity of VLRs that result from this combinatorial assembly are comparable to those of antigen receptors in jawed vertebrates (Alder *et al.*, 2005).

RNA interference. RNA interference (RNAi) has an important role in antiviral immunity in plants and invertebrates (Hamilton and Baulcombe, 1999; Li, Li and Ding, 2002; Lu *et al.*, 2005; van Rij *et al.*, 2006; Wang *et al.*, 2006), and recent studies proposed a role in mouse embryonic stem cells, which lack a functional interferon response (Li *et al.*, 2013; Maillard *et al.*, 2013). In antiviral RNAi, an RNaseIII enzyme known as Dicer recognizes and cleaves viral double-stranded RNA into viral small interfering RNA (viRNA) that is typically 21–24 nucleotides in length (Fig. I/2b). These viRNAs are loaded onto Argonaute enzymes, which are a key component of the RNA-induced silencing complex (RISC), to guide the binding and cleavage of complementary viral RNA (reviewed in Ding and Voinnet, 2007). The propensity to generate a diversity of viRNAs is encoded by the host RNAi genes, but generating resistance-allele diversity does not require any genetic change to the host genome. The sampling of viRNAs from viral RNA is non-random (Molnar *et al.*, 2005), but it leads to viRNA diversity within and between individuals. In the absence of such diversity of viRNAs, rapid virus evolution can overcome RNAi-mediated resistance through mutations of the viral target sequence that disrupt complementarity to the viRNA (Lafforgue *et al.*, 2011; Holz *et al.*, 2012; Martinez *et al.*, 2012). As a consequence of the rapid escape of viruses from a single viRNA, the mainstream approach to developing transgenic crops with RNAi-mediated virus resistance depends on the production of multiple viRNAs (Swaney *et al.*, 1995).

CRISPR–Cas. Although some prokaryotes do encode Argonaute enzymes (Swarts *et al.*, 2014), the main prokaryotic DGM is the CRISPR–Cas adaptive immune system (Fig. I/2c). CRISPR–Cas systems integrate parasitic DNA sequences (spacers) into CRISPR loci on the host genome (Barrangou *et al.*, 2007), which provides heritable immunity against pathogens with a matching nucleotide sequence. Cas enzymes use processed transcripts of CRISPR loci as guides (Brouns *et al.*, 2008) to mediate the sequence-specific cleavage of (usually) complementary DNA (Garneau *et al.*, 2010) and sometimes RNA (Abudayyeh *et al.*, 2016; East-Seletsky *et al.*, 2016) or both (Goldberg *et al.*, 2014; Samai *et al.*,

2015; Jiang *et al.*, 2016). Although the sampling of pathogen-derived spacers is often not random (Richter *et al.*, 2014; Modell, Jiang and Marraffini, 2017), most bacterial clones acquire unique spacers (Westra *et al.*, 2015). Hence, CRISPR-Cas systems rapidly generate population-level diversity at CRISPR loci. One study showed that this diversity is an important fitness determinant, as it limits the evolution of CRISPR escape phage, which carry point mutations that allow evasion from the CRISPR-mediated immune response (van Houte *et al.*, 2016). CRISPR–Cas systems can also generate within-host diversity through the acquisition of multiple spacers from the same pathogen, which also limits the evolution of escape phage (Levin *et al.*, 2013; van Houte *et al.*, 2016).

Phase variation. Another example of a bacterial DGM is phase variation, which can generate population-level heterogeneity by switching genes on or off through slipped-strand mispairing, inversions or site-specific recombinations (Bikard and Marraffini, 2012). A key example in bacterial immunity is provided by type I restriction–modification (RM) shufflons (Fig. I/2d). RM systems function through epigenetic modification of specific host DNA sequences and cleavage of the same sequences when unmodified, such as those encoded on phage genomes (reviewed in van Houte, Buckling and Westra, 2016). Type I RM shufflons encode specificity subunits (HsdS), which determine the sequence specificity; methylases (HsdM), which catalyse methylation of the corresponding sequences; and restriction endonucleases (HsdR), which catalyse cleavage of these sequences when unmodified. Phage can rapidly evolve to overcome RM systems, for example when HsdM modifies phage sequences before cleavage by HsdR (Levin, Antonovics and Sharma, 1988; Korona and Levin, 1993), which enables unchallenged spread of the methylated escape phage through the bacterial population. However, HsdS in type I RM shufflons can rapidly diversify through recombination between multiple *hsdS* genes, hence generating population-level diversity in RM specificity (Dybvig, Sitaraman and French, 1998; Tettelin *et al.*, 2001). These systems often generate diversity between hosts, but sometimes, when one host encodes multiple HsdS at multiple loci that can be active simultaneously (Dybvig, Sitaraman and French, 1998), they can also generate within-host diversity. Both levels of diversity are predicted to limit the evolution and spread of RM escape phage (Sneppen *et al.*, 2015).

Unifying concepts. As outlined above, targeted DGMs tend to generate high levels of diversity, often both between and within hosts. This diversity is exclusively generated in genes involved in host–pathogen interactions, and therefore, these DGMs do not usually provide benefits in the absence of pathogens. However, exceptions in which these DGMs have been co-opted for other processes do exist. For example, RNAi is involved in many processes including gene regulation, epigenetic modification and transposon regulation, and it therefore has a crucial role in organismal homeostasis and phenotypic plasticity (Ketting, 2011). Bacterial CRISPR-Cas systems sometimes also have a role in (virulence) gene regulation (Westra, Buckling and Fineran, 2014), such as bacterial lipoprotein expression by *Francisella novicida* during infection to avoid triggering a host immune response (Sampson *et al.*, 2013). Bacterial RM systems can regulate gene expression through epigenetic changes, such as capsule gene expression in *Streptococcus pneumoniae*, which is a key virulence determinant (Manso *et al.*, 2014). However, targeted DGMs usually have specialized roles in immunity, and where they have non-immune functions, their role in immunity may be less important (Seo *et al.*, 2013).

Benefits of DGMs

Most DGMs are likely to have evolved because they increase an individual's resistance to pathogens. For example, sexual reproduction can generate rare genotypes that can escape infection with common pathogens (Hamilton, Axelrod and Tanese, 1990), and increased mutation rates in bacteria increase the rate at which mutations that confer resistance to phage are generated (Pal *et al.*, 2007; Morgan, Bonsall and Buckling, 2010). In these cases, DGM-associated genes can increase in frequency together with the beneficial alleles or allele combinations they generate owing to genetic linkage; for example, bacterial mutator alleles, such as mutations in the methyl-directed mismatch repair system, can be selected for based on their linkage with beneficial phage resistance mutations that are generated in the same genetic background; these resistance mutations occur at higher frequencies in mutator strains compared with wild-type bacterial strains because of the increased mutation rate (Taddei *et al.*, 1997). Other systems, such as RNAi and the adaptive immune response, are also likely to have

arisen because of their benefits in terms of an individual's resistance to pathogens, and the within-host diversity that they generate will reduce the probability of pathogens overcoming host defences.

However, theory and data show that host diversity (and therefore DGMs, provided that the diversity that they generate is maintained) can also provide population-level benefits. Increased host diversity can decrease the pathogen reproductive rate (epidemiological effect) and can make it more difficult for the pathogen to adapt to the host population (evolutionary effect). As targeted DGMs usually generate higher levels of diversity in resistance genes than do non-targeted DGMs, they provide greater epidemiological and evolutionary benefits in the presence of infectious diseases.

Epidemiological effect. The first population-level benefit of host diversity is a reduction in the size of disease epidemics. If host–pathogen interactions are specific (in other words, pathogen genotypes infect a restricted and non-overlapping range of host genotypes), then the reproductive rate of a pathogen is predicted to be larger in homogeneous compared with heterogeneous host populations (Lively, 2010b). This epidemiological effect results from a reduction in the frequency of productive infections (in other words, the host genotypes that can be infected by a particular pathogen genotype become diluted amidst other host genotypes that are not susceptible to infection, which is known as a dilution effect (Keesing *et al.*, 2010)), as well as an increase in the frequency of failed infections when pathogens infect resistant host genotypes. Above a threshold level of host diversity, the reproductive rate of the pathogen becomes smaller than 1, and it will become extinct unless it can evolve an altered host range.

Evolutionary effect. The diversity of host-resistance alleles is also predicted to limit pathogen evolution because it is more difficult for pathogens to adapt to heterogeneous host populations than to monocultures (Hamilton, Axelrod and Tanese, 1990) and because the reduction in pathogen reproductive rate in heterogeneous host populations reduces the evolutionary potential for the pathogen to adapt to the host population (Antia *et al.*, 2003). For example, a field study carried out in China that examined the effect of mixing pathogen-resistant and -susceptible rice crops found that mixed crops had a greatly decreased severity of rice blast compared with monocultures, and the results suggested that there

was reduced pathogen adaptation to the plants' resistance alleles in the long term when the crops were mixed (Zhu *et al.*, 2000). A laboratory evolution study showed that population-level diversity generated by CRISPR–Cas adaptive immune systems limits the evolution of CRISPR escape phage. When bacterial hosts were grown in monoculture, CRISPR escape phage emerged rapidly, but when the same bacterial CRISPR clones were mixed, escape phage were never observed and the phage were driven to extinction (van Houte *et al.*, 2016). Diversity in terms of antibodies and viRNAs is likely to be associated with similar benefits — in these cases applying to both individuals (owing to within-host diversity) and populations (owing to between-host diversity).

Costs of DGMs

DGMs may be selected against if they carry fitness costs that outweigh their benefits. These costs will differ between DGMs and may be particularly high for untargeted DGMs. For example, most random mutations are either neutral or detrimental, being the underlying cause of various diseases (Forsberg, Gisselsson and Dumanski, 2017; Tubbs and Nussenzweig, 2017). However, the costs associated with high mutation rates can be alleviated through mechanisms for tissue-specific mutation (for example, in B cells), site-specific mutation (for example, somatic hypermutation of antigen receptor genes) or stress-induced mutation (as occurs in some bacteria) (Foster, 2007). Sexual reproduction is also associated with a large cost compared with parthenogenesis, the most important cost being the reduced number of offspring per adult, as males do not contribute to reproductive output during sexual reproduction (known as the twofold cost of sex) (Maynard Smith, 1971, 1978).

Targeted DGMs can be associated with both immunopathological costs, whereby the immune response damages the host, and energetic costs. These costs can be manifested both in the short term (for example, when immunological activity contributes to or worsens disease symptoms (reviewed in Graham, Allen and Read, (2005))) and in the long term (for example, by decreasing the lifespan of individuals with strong immune responses (Finch and Crimmins, 2004)). Minimizing these costs has probably had an important role in shaping the evolution of immune systems (reviewed in Graham, Allen and Read (2005)). For example, the production and disposal of large numbers of B and T cells owing to non-productive gene rearrangements during V(D)J recombination and CSR impose an

energetic cost on the individual, which has led to the evolution of specific mechanisms to reduce the frequency of deleterious conformations, such as orientation-specific joining of gene fragments in the same transcriptional orientation (Dong *et al.*, 2015). In the case of CRISPR–Cas, the immune response is associated with a fitness cost that is induced upon infection, which may be an energetic cost as a result of induced expression of CRISPR–Cas or an immunopathological cost owing to self-reactivity, or it could result from pathogen-induced damage to the host before pathogen clearance (Vale *et al.*, 2015; Westra *et al.*, 2015). Fitness costs resulting from immunopathology are frequently observed in the context of vertebrate immune systems (Kobasa *et al.*, 2007; Graham *et al.*, 2010), as well as bacterial innate (Pleška *et al.*, 2016) and adaptive (Stern *et al.*, 2010) immune systems. One example of immunopathology associated with the antibody response is that which is mediated by RAG1. During V(D)J recombination, RAG1 often binds cryptic recombination sequences in non-immunoglobulin sequences, leading to genome instability; the cost of this is mitigated by reducing the genome-wide frequency of such sites (Teng *et al.*, 2015).

Whether the benefits of a particular DGM outweigh the costs will depend on the environment, most notably pathogen density. In the cases of mutation and sexual reproduction, the fitness costs are constitutive, and hence, these DGMs may be selected against in the absence of pathogens (Ashby and King, 2015). This is consistent with the observation that sexually reproducing snails are more abundant in environments with high pathogen densities (discussed above). By contrast, targeted DGMs tend to have relatively low fitness costs in the absence of pathogens and are therefore less likely to be strongly selected against when pathogens are absent (Westra *et al.*, 2015).

DGM-associated co-evolution

The individual- and population-level benefits of DGMs indicate that pathogens that are unable to evolve to overcome host resistance owing to genetic constraints, such as their smaller genome size (Morgan, Gandon and Buckling, 2005), may become extinct. For example, during co-evolution between the bacterium *Pseudomonas fluorescens* and its phage under laboratory conditions, host resistance and pathogen infectivity increase over time in a process known as arms-race dynamics (ARD). This process is associated with selective sweeps of

host and pathogen genotypes with increased resistance and infectivity, which become rapidly fixed in the population (Buckling and Rainey, 2002), resulting in the maintenance of low levels of host and pathogen diversity. An increase in the bacterial mutation rate resulting from mutations in the methyl-directed mismatch repair system allows bacteria to outpace the phage, as it becomes increasingly difficult for the phage to adapt to the most common but rapidly evolving host genotype (Morgan, Bonsall and Buckling, 2010), resulting in an increased probability of phage extinction (Pal *et al.*, 2007).

However, in other cases, the diversity of host resistance alleles and pathogen infectivity alleles is maintained in the population, and host resistance can be quite specific for particular pathogen genotypes (Luijckx *et al.*, 2011, 2013). Indeed, this assumption is crucial to explain pathogen-mediated maintenance of sexual reproduction (Hamilton, Axelrod and Tanese, 1990; Lively, 2010a). In this case, pathogens would typically adapt to counter resistance of the most common host genotype (Lively and Dybdahl, 2000). This would provide an advantage to rare host genotypes (Lively, Craddock and Vrijenhoek, 1990), which theoretically leads to ongoing co-evolution between the host and pathogen through fluctuating selection dynamics (FSD), whereby fitness and therefore frequencies of host and pathogen genotypes fluctuate over time (Hamilton, Axelrod and Tanese, 1990). Co-evolutionary FSDs have been observed in several host–pathogen interactions (Decaestecker *et al.*, 2007; Gomez and Buckling, 2011). Unlike ARD, FSD-type co-evolution can in theory continue indefinitely, which allows for the maintenance of host and pathogen diversity.

However, not all DGMs that generate high levels of diversity will cause FSD-type co-evolution. In addition to the requirement for a high specificity of infection (discussed above), pathogens need to be able to adapt to a common host genotype for FSD to occur. This depends both on the molecular mechanism of the DGM (for example, genetic constraints on the pathogen to overcome host resistance will be different for an antibody response, where pathogens need to evolve a modified epitope, compared with RNAi or CRISPR–Cas, where a SNP in the target sequence leads to escape) and on the levels of host diversity that are generated. In the case of untargeted DGMs, which generate relatively low levels of diversity, pathogen adaptations by mutation or recombination alone are often sufficient to overcome host resistance. As discussed above, the high levels of diver-

sity generated by targeted DGMs can limit pathogen adaptation by mutation, resulting in pathogen extinction (van Houte *et al.*, 2016). For example, despite the rapid evolution of CRISPR escape phage (Deveau *et al.*, 2008) and the high specificity of interactions between CRISPR-mediated resistant hosts and escape phage (one escape phage can infect only a single CRISPR-mediated resistant clone (Morley *et al.*, 2017)), FSD-type co-evolution is not observed owing to the high levels of CRISPR diversity (Chabas *et al.*, 2016; van Houte *et al.*, 2016), which limit the ability of phage to adapt to dominant host genotypes. This relationship, in turn, imposes a strong selection pressure on pathogens to evolve more sophisticated strategies to evade host immunity. As the evolution of such anti-DGM strategies requires genetic innovation, this process occurs over longer timescales as part of an arms race between hosts and pathogens.

Evolution of anti-DGM strategies. One way in which pathogens can escape the host resistance mediated by targeted DGMs that generate high levels of diversity is by evolving high levels of diversity themselves. Probably the best-known example of this is capsule switching in the human pathogen *Streptococcus pneumoniae*, which allows the pathogen to evade the human immune system during infection (Griffith, 1928). Another example is antigenic variation in *Trypanosoma brucei*, the causative agent of sleeping sickness, which has approximately 1,000 copies of the variant-specific glycoprotein (VSG) gene that encodes the main surface antigen (Fig. I/3a). Each pathogen expresses only one VSG gene and the immune response targets the most abundant VSG genotype, which allows rarer genotypes to evade the immune response (Schwede *et al.*, 2015). Similarly, *Plasmodium falciparum*, the human malaria pathogen, encodes a large number of *var* genes, which are differentially expressed in different pathogen lines. These genes encode surface proteins responsible for antigen variation in infected red blood cells, which are key determinants of immune evasion and virulence (Su *et al.*, 1995). Other examples include *influenza* virus reassortment (Gerber *et al.*, 2014) (Fig. I/3b) and diversity-generating retroelements found in the tail genes of phage (Doulatov *et al.*, 2004; Paul *et al.*, 2015) (Fig. I/3c). In the case of the nematode *Strongyloides ratti*, pathogen sexual reproduction helps the pathogen to adapt to the host immune system; compared with wild-type rats, the infection of immunocompromised rats is characterized by a decreased proportion of facultatively sexual *S. ratti* (Gemmill, Viney and Read, 1997).

A second common strategy by which pathogens can escape host resistance is through the evolution of pathogen-encoded mechanisms that specifically antagonize host DGMs. For example, anti-DGM strategies seem to be nearly ubiquitous in plant viruses, which almost invariably encode antiviral silencing suppression genes that interfere with the RNAi pathway (reviewed in Pumplin and Voinnet (2013) (Fig. I/3d). Many vertebrate parasites and pathogens also encode proteins that block the immune response (Belkaid *et al.*, 2002; Young, Hussell

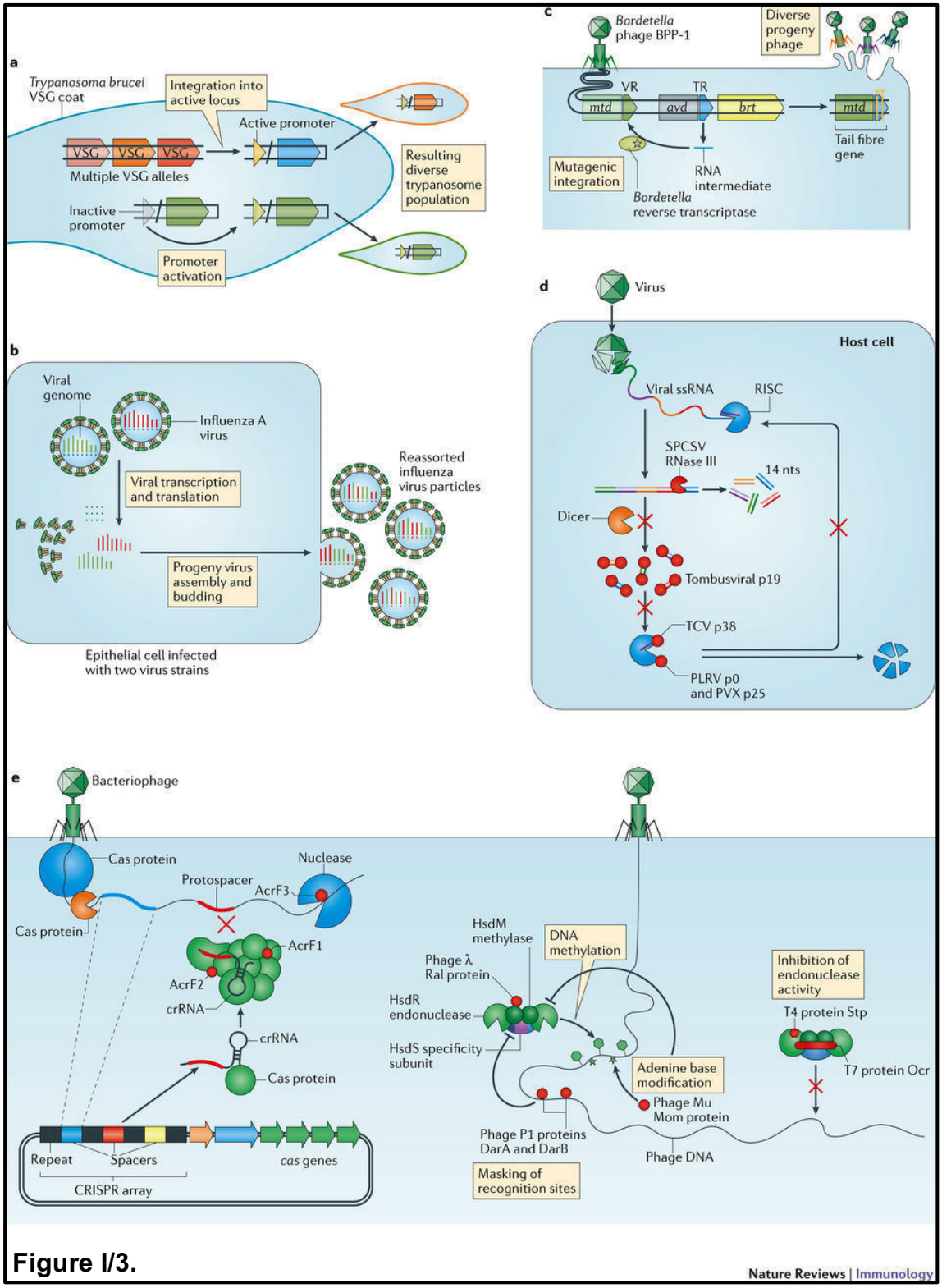


Figure I/3.

Figure I/3. Anti-diversity-generating mechanism strategies.

a *VSG* encodes the variable surface glycoprotein (VSG) coat of Trypanosoma brucei. *VSG* alleles are encoded in the *VSG* array, on mini-chromosomes (not shown), and in alternative *VSG* loci. Translocation of alleles into the single active locus or activation of an alternative promoter causes population-level diversity in VSG expression. **b** Genetically distinct strains of *influenza* virus infect an epithelial cell. During progeny virus assembly, RNA strands reassort to create diverse *influenza* virus particles. **c** Some phage encode diversity-generating retroelements for diverse tail fibre genes. An invariable template repeat (TR) is copied into a variable repeat (VR) within the major tropism determinant (*mtd*) gene. This process of mutagenic homing requires an RNA intermediate and a unique low-fidelity reverse transcriptase. *avd*, accessory variability determinant (cofactor for *brt*); *brt*, *Bordetella* reverse transcriptase. **d** Many plant virus proteins affect RNA interference (RNAi). Sweet potato chlorotic stunt virus (SPCSV) RNase III cleaves viral dsRNA into 14-nucleotide (nt)-long segments that are inaccessible to the RNA-induced silencing complex (RISC). Tombusviral p19 achieves this by binding virus-derived small interfering RNA (viRNA) molecules. Turnip crinkle virus (TCV) p38 binds Argonaut proteins to block viRNA loading. Polerovirus (PLRV) p0 and potato virus X (PVX) p25 promote degradation of RISC. **e** Bacteriophage anti-CRISPR (Acr) proteins inhibit CRISPR–Cas (left side of panel). AcrF1 and AcrF2 interact with the Cascade complex to prevent it from binding complementary DNA; AcrF3 blocks recruitment of Cas nuclease to Cascade, which prevents DNA cleavage. Phage anti-restriction–modification (RM) proteins inhibit phage restriction (right side of panel). Phage λ protein Ral stimulates methylase (HsdM) to methylate phage DNA, phage Mu-encoded Mom protein modifies adenine bases in phage DNA, and phage P1 proteins DarA and DarB bind to restriction sites; these three mechanisms protect phage DNA from degradation. T7 protein Ocr binds to the restriction endonuclease and stops it from binding DNA. T4 protein Stp changes the conformation of the endonuclease to disrupt its activity.

and Dougan, 2002), and many bacteriophage encode anti-RM (Atanasiu *et al.*, 2002) and anti-CRISPR proteins (Bondy-Denomy *et al.*, 2013) (Fig. I/3e). The latter were first discovered in *Pseudomonas* phage (Bondy-Denomy *et al.*, 2013)

but are now recognized to be extremely widespread, to have little sequence similarity (Pawluk, Staals, *et al.*, 2016) and to be mechanistically diverse (Bondy-Denomy *et al.*, 2015; Pawluk, Amrani, *et al.*, 2016; Rauch *et al.*, 2017).

The evolution of anti-DGM strategies by pathogens can result in an arms race with the host that takes place over long evolutionary timescales, whereby hosts continuously evolve to escape anti-DGM activity and pathogens evolve to increase anti-DGM activity. Indeed, genes that are involved in antiviral RNAi are among the most rapidly evolving genes in the *Drosophila* genome (Obbard *et al.*, 2006). Such co-evolutionary interactions are likely to have an important role in shaping immune systems, including their genetic and functional diversification.

Conclusion and future prospects

Many hosts encode DGMs that are involved in host–pathogen interactions. Here, we propose that although their molecular mechanisms are diverse, many DGMs have common features with regards to their selective benefits and epidemiological and co-evolutionary consequences. Understanding the link between these microscopic and macroscopic processes will be important for designing effective public-health strategies to prevent or limit the emergence of infectious diseases. Vaccination programmes and crop protection strategies in agriculture rely on increasing the proportion of resistant hosts in the population, but they often do not take into account the diversity of resistance alleles. Most of our current understanding of the interaction between diversity and disease relies on theoretical models, but many of these models have not been experimentally validated. Future work aimed at understanding how environmental variables such as pathogen diversity and population structure impact the benefits of host diversity will be an important step forward in our understanding of the evolutionary epidemiology of pathogens, which can help to develop durable strategies to control human, plant and animal disease. Understanding the epidemiological and evolutionary implications of resistance-allele diversity may also guide further molecular studies to improve our understanding of how diversity in resistance alleles, such as antibodies and small RNA molecules, is generated and the scale at which diversity is maintained over time both within and between individuals.

Chapter II: Restriction-Modification shufflons are a means of generating population-level diversity

Abstract

Restriction-Modification (RM) is the most widespread bacterial immune system and protects bacteria from bacteriophage (phage) by cleaving foreign DNA sequences, while methylating its own genome to avoid self-recognition. In type I RM systems, specificity subunit HsdS confers DNA sequence specificity when associating with methyltransferase and restriction endonuclease. Because phage can escape from this response by becoming methylated themselves with relative ease, it is unclear whether RM provides a large benefit when bacteria coevolve with phage over a longer period. Many bacteria have evolved mechanisms to obtain novel RM recognition sequence specificities which could improve the benefits of RM to bacteria. Here, by means of a combination of bioinformatics, *in silico*, and literature analyses, we estimate that at least 8% of Type I RM systems have the capacity to switch specificity. The simplest of these mechanisms involves a switch in expression of alternative *hsdS* genes, which encode the sequence specificity subunit of the RM system. Others acquire novel *hsdS* sequences through plasmids, and the more sophisticated shufflons recombine multiple *hsdS* pseudogenes to rapidly generate new specificities. *Mycoplasma pulmonis* has two RM-encoding *hsd* loci with two copies of *hsdS* containing several recombination sites each. Analysing this Mpu shufflon *in silico* revealed a repertoire of 30 different possible recombinations, with 12-14 of these specificities likely to be active *in vivo* as they encode the traditional number of 1-2 target recognition domains. Shufflons such as this are likely to drastically change the coevolutionary dynamics of bacteria and phage, providing a large benefit to bacteria through generation of population-level diversity.

Introduction

As discussed in the previous chapter, diversity-generating mechanisms are integral to host-parasite interactions and can be greatly beneficial to hosts in the face of an epidemic (Westra *et al.*, 2017). The bacterial adaptive immune response CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated) can generate population-level diversity of bacteriophage

(phage) recognition and cleavage (van Houte *et al.*, 2016). The more abundant innate bacterial immune response of Restriction-Modification (RM) typically only recognises predetermined DNA sequences and therefore does not confer the same benefits. As a nearly ubiquitous immune mechanism (about 90% of bacteria and archaea encode one or more RM systems (Stern and Sorek, 2011)), RM is nonetheless an important barrier to horizontal gene transfer and plays a role in many coevolutionary dynamics of bacteria with phage. However, in some cases RM diversity can be generated too.

RM systems utilise methylation of the bacterial genome as a simple means of self/non-self discrimination to cleave unmethylated foreign DNA. When invading a bacterium, a phage's DNA can accidentally become methylated by RM. This protects the phage and its progeny from cleavage and allows them to become invisible to the immune response. Due to the relative ease with which phage can escape, it has been argued that RM systems only provide a transient advantage to bacteria when invading a different ecosystem with new phage (Korona and Levin, 1993). These dynamics are likely to change when an RM system has the ability to change its specificity, which is highly beneficial to bacteria: a bacterial population with a reservoir of different RM specificities is protected from an escape phage sweeping through the entire population.

Mechanistically there are four types of RM systems, most of which have a methyltransferase and a restriction endonuclease to achieve self/non-self recognition and cleavage of invasive DNA. The methyltransferase modifies bacterial DNA at a specific sequence, while the restriction endonuclease recognises the same sequence and catalyses cleavage when it is unmethylated. Type IV systems are the exception to this rule, these lack a methyltransferase and cleave modified DNA while leaving the bacterial genome unmethylated (Ershova *et al.*, 2015). Both Type II and III systems are comprised of only two genes, *mod* and *res*, coding for these two essential components. In these systems, specificity for the same DNA sequence is separately defined in both genes (Type II), or in the methyltransferase only (Type III). In contrast, Type I RM systems encode a specificity subunit that associates with the methylase and restriction endonuclease. The Type I *hsd* (host specificity for DNA) locus encodes this specificity subunit HsdS (S) and HsdM (M) (Arber and Linn, 1969), which together form a functional methyltransferase complex with a M_2S_1 stoichiometry that methylates target DNA sequences. The last *hsd*-encoded component, the restriction subunit HsdR (R),

can further associate with the other proteins to form the $R_2M_2S_1$ complex that has the ability to both methylate and cleave unmodified target sequences (Burckhardt *et al.*, 1981; Dryden *et al.*, 1997). Upon encountering an unmethylated target sequence, both HsdR subunits in this complex will begin to ATP-dependently translocate DNA, until one encounters another activated HsdR subunit and cleavage is catalysed (Janscak *et al.*, 1999). Due to this requirement, DNA cleavage can only be catalysed if at least two unmodified target sequences are present on the same DNA strand and cleavage occurs at a random position in between the two targets. Target sequences consist of two sequences of 3-4 nucleotides separated by a stretch of several, often 5-8 unspecific nucleotides (e.g. EcoKI's recognition sequence of 5'-AACNNNNNNGTGC-3' (Kan *et al.*, 1979)). This is due to the structure of HsdS, which contains two DNA-binding target recognition domains (TRDs) separated by a linker domain (Fig II/1). This structure allows different kinds of mutations of *hsdS* (either within the TRDs to alter recognition sequence, or within the linker sequence to alter spacing between them) to have a great impact on DNA specificity. Compared to Type II or III RM systems, *hsdS* and its structure allow for rapid evolution of new diversity in Type I systems, because a single gene which has large capacity to be altered confers specificity to the entire RM system. Therefore, while diversity-generating mechanisms can be

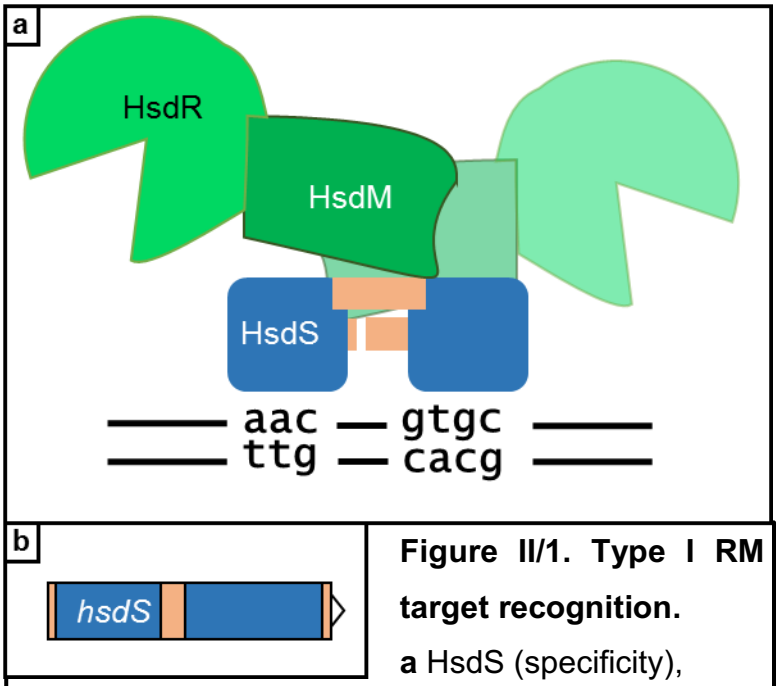


Figure II/1. Type I RM target recognition. **a** HsdS (specificity), HsdM (methyltransferase), and HsdR (restriction endonuclease) subunits are the three components of the RM protein complex. Two subunits each of HsdM and HsdR associate with HsdS. HsdS consists of two target recognition domains (blue) which bind specific DNA sequences and two spacer domains (beige) which separate the TRDs. **b** Structure of a typical *hsdS* gene. Beige regions get translated as spacer domains, blue regions as TRDs. Figure adapted from Murray (2000); Fig. 2.

found in Type III RM systems too (e.g. in *Helicobacter pylori* (Kojima *et al.*, 2016)), the capacity for these is far greater for Type I systems (Murray, 2000).
In this chapter, I perform a bioinformatics analysis to understand how common diversity generating Type I RM systems are, and an *in silico* analysis of one such RM shufflon from *Mycoplasma pulmonis* to examine its potential for diversity generation.

In silico analysis, Results, and Discussion

The ability to generate population-level diversity is likely a key benefit of Type I RM shufflons, whose unique architecture allows for rapid evolution of new RM specificities. In order to understand how widespread these diversity-generating mechanisms might be, we conducted a search for Type I RM systems with multiple *hsdS* genes listed on REBASE (Roberts *et al.*, 2015) as of September 2016. While this search did not reveal RM systems in which S subunits have different names (e.g. *M. pulmonis* Mpu shufflon, see below), and may include some non-functional RM systems that have lost *hsdR*, it nevertheless provides an overview of how many RM systems have a reservoir of *hsdS* genes as a basis of diversity. We found that 1117 (8%) Type I RM systems encode at least two different specificity subunits. Of these, only 18 encode two methylases, indicating that the vast majority are single RM systems with multiple specificities. 295 (2%) encode at least three, and 69 (<1%) encode four or more HsdS subunits. Two RM systems, both found in different *Myoplasma haemofelis* strains, encode 13 different *hsdS* variants. Notably, other RM systems with high numbers of *hsdS* genes (7-9 variants) are also encoded by *M. haemofelis* strains. Specifically, *M. haemofelis* strains were found to have up to 21 *hsdS* genes, not all of which are functional (Santos *et al.*, 2011). RM systems with 6 *hsdS* genes can be found in *Enterococcus cecorum* and *Mycoplasma capricolum*. In *Mycoplasma* species, presence of multiple S subunits points towards regulation of their expression by inversions for phase variation (Brocchi, de Vasconcelos and Zaha, 2007). However, often this can be associated with regulation of gene expression rather than phage defense. Overall, this data shows that at least 8% of Type I RM systems have the capacity for either switching, or even generating new specificity.
Literature research and more detailed analysis of the identified Type I RM DGMs reveals two main ways to generate diversity: horizontal transfer of novel specificity subunits and recombination (Fig II/2).

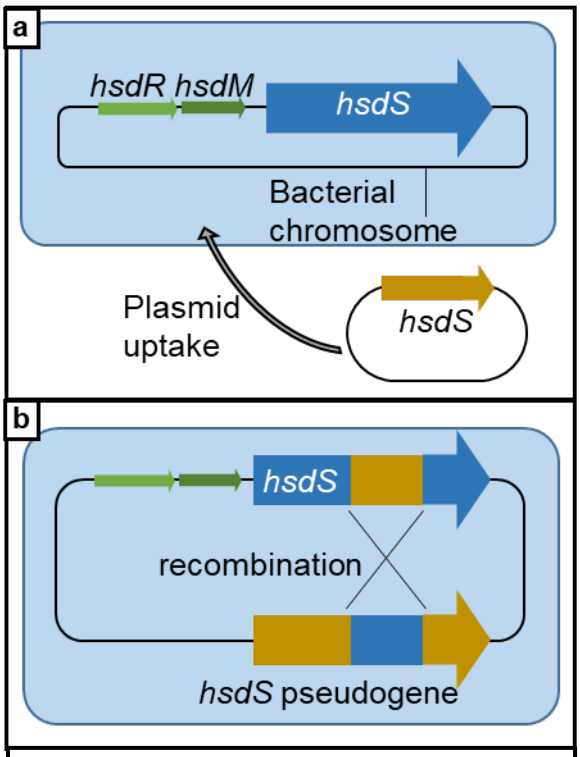


Figure II/2. Diversity-Generating Type I RM Mechanisms.

a Bacteria such as *Lactococcus lactis* can horizontally acquire *hsdS* plasmids, which confer a novel RM specificity. **b** Recombination can occur with chromosomally encoded *hsdS* pseudogenes or with *hsdS* genes encoded on plasmids (not shown). This can lead to altered phage resistance or epigenetic phase variation.

For example, some *Lactococcus lactis* strains chromosomally encode a Type I RM system which can acquire new specificities if the bacteria take up one of several *hsdS* plasmids that can be transferred through a population (Schouler *et al.*, 1998). Plasmid-encoded RM systems in other *L. lactis* strains can acquire novel specificity through acquiring *hsdS* plasmids or through recombination between *hsdS* genes on different plasmids (O'Sullivan *et al.*, 2000). *Salmonella typhimurium hsdS* genes have been found to recombine with a *Salmonella potsdam hsdS* gene to generate a novel specificity (Gann *et al.*, 1987). Taking basic recombination to the extreme, shufflon RM systems generally utilise short inverted repeats within *hsdS* to recombine two or several *hsdS* pseudogenes, generating novel specificity. Few examples of shufflon RM systems are known, but they are broadly distributed across different bacterial species. In *Bacteroides fragilis* strains, inverted repeats within *hsdS* of several closely related shufflons facilitate TRD rearrangements. Shufflon BB in *B. fragilis* NCTC 9343 has the capacity for eight HsdS variants (Patrick *et al.*, 2010). A *Streptococcus pneumonia* strain was found to encode two additional *hsdS* pseudogenes which recombine to create up to four novel specificities (Tettelin *et al.*, 2001). A related *S. pneumonia* strain encodes a shufflon that can generate up to six specificities, however in both strains a switch in RM specificity is associated with phase variation through differential gene expression by methylation patterns rather than phage resistance (Manso *et al.*, 2014). For the *M. pulmonis* Mpu shufflon, eight *hsdS*

variants have been characterised, although this shufflon with multiple recombination sites has an even larger capacity for novel specificities (Dybvig, Sitaraman and French, 1998). As the *M. pulmonis* shufflon has unprecedented capacity for recombination and the differential phage restriction capabilities of its various specificities are well documented, we will analyse this Mpu shufflon in detail for the remainder of this chapter. A greater understanding of this diversity-generating mechanism will allow the creation of a model system in which we can study phage-bacteria interactions in the presence of multiple RM specificities as well as the interactions of RM with other bacterial immune systems.

M. pulmonis is a Firmicute murine pathogen implicated in airway infections. Closely related *Mycoplasma* species, such as *M. pneumonia*, are implicated in

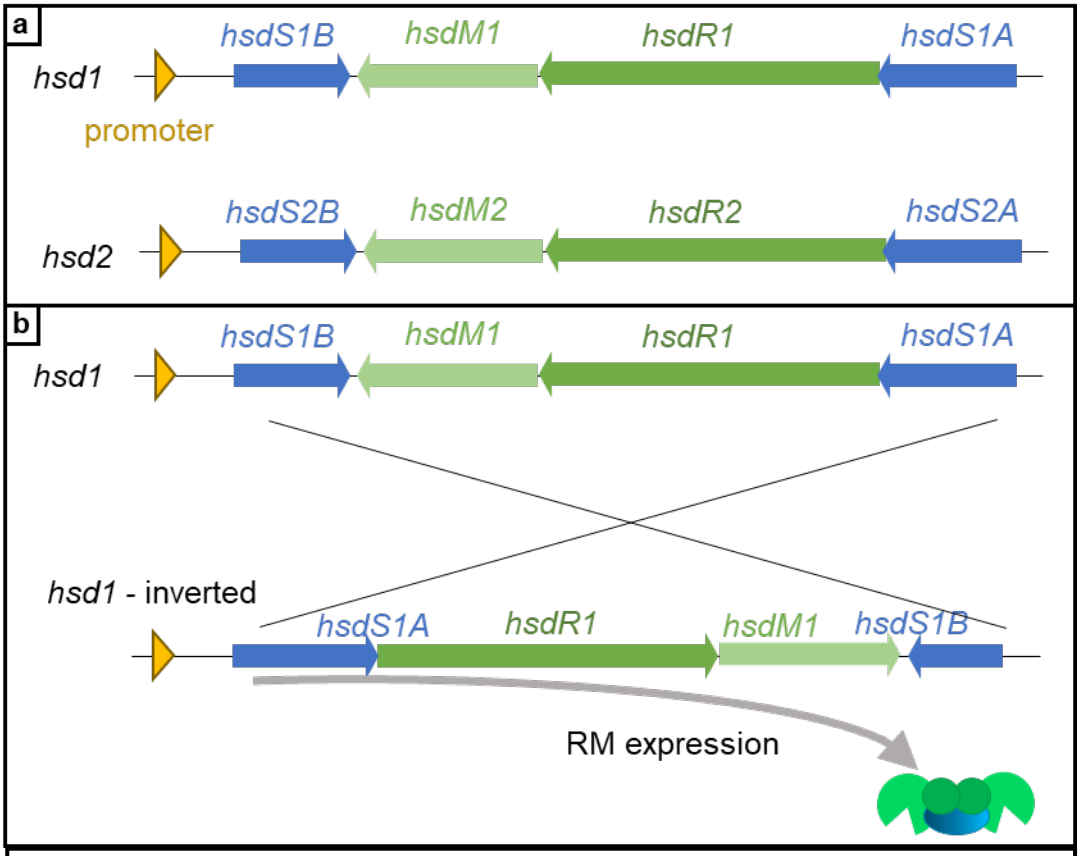


Figure II/3. The Mpu shufflon as an ON/OFF switch.

a Layout of the *Mycoplasma pulmonis hsd* loci with genes drawn to scale. **b** A single inversion event using *hsdS* recombination sites leads to RM expression where all genes can be transcribed. This is the same for *hsd2* (not shown).

human infections. While a switch in HsdS expression in *M. pulmonis* can be associated to a certain extent with the bacterium's infection stage in rats (Gumulak-

Smith *et al.*, 2001), this is far from its only role: a different restriction profile renders a bacterial host immune to *Mycoplasma* phage amplified on a host with a different RM specificity. The Mpu shufflon consists of two *hsd* loci, both encoding nearly identical HsdR and HsdM subunits (Fig II/3a). Each locus contains one copy of *hsdS* as well as one *hsdS* pseudogene (Sitaraman and Dybvig, 1997). Short *vip* (vipareetus) and *hrs* (*hsd* recombination site) recombination sequences within *hsdS* allow recombination with matching sequences on the opposing pseudogene of the same locus (Dybvig, Sitaraman and French, 1998). Due to the directionality of Mpu *hsd* operons, HsdR and HsdM are only expressed after an inversion has occurred in either locus (Fig II/3b), for this reason the Mpu shufflon had initially been discovered as a RM on/off switch regulated by inversions (Dybvig and Yu, 1994). Recombination between *vip* and *hrs* sites is catalysed by a recombinase which curiously also triggers genetically and phenotypically unrelated *vsa* (variable surface antigen) recombination in *M. pulmonis* (Sitaraman, Denison and Dybvig, 2002). While all *vip* sequences within *hsdS* pseudogenes are identical, some *hrs* sequences contain a single nucleotide insertion which is not sufficient to prevent recombination. Due to the similarity of *hsdM* and *hsdR* between the loci, an HsdS subunit can associate with HsdR and HsdM from either locus to form a functional $R_2M_2S_1$ complex.

hsdS genes in both loci are composed of 3-4 gene segments, which we defined as the sequence between two adjacent recombination sites (Fig II/4). Three *hsd2* gene segments (red, yellow, violet) are identical to their homologues in *hsd1*. Grey and orange segments are nearly identical, with only 7 and 2 nucleotide substitutions corresponding to 3 and 2 amino acid changes between the loci respectively (Table II/1). The only gene segment unique to *hsd2* (green) shares 98% homology with the grey segment in the final ¾ of its sequence. Therefore, disregarding the minor differences in grey and orange segments as we will do from here on, there is a reservoir of 7 unique gene segments for *hsdS* variants, whereby grey is always the first segment within a sequence.

To determine the capacity for *hsdS* diversity in the Mpu shufflon, we simulated inversions between homologous recombination sequences *in silico* and determined all unique *hsdS* sequences that could arise through single, double, or triple inversions. Any additional inversions would revert the gene back to a sequence that can be produced by fewer recombination events. We found that the Mpu shufflon can produce *hsdS* genes with 30 different sequences (Table II/2), 7 of

which (MpuUI-VI, MpuUVIII) had previously been sequenced and tested against phage by Dybvig, Sitaraman and French, (1998). Subunits not previously defined were named MpuT1-29 (T for theoretical), adhering to the convention of odd-numbered subunits being produced by *hsd1* and even-numbered subunits being

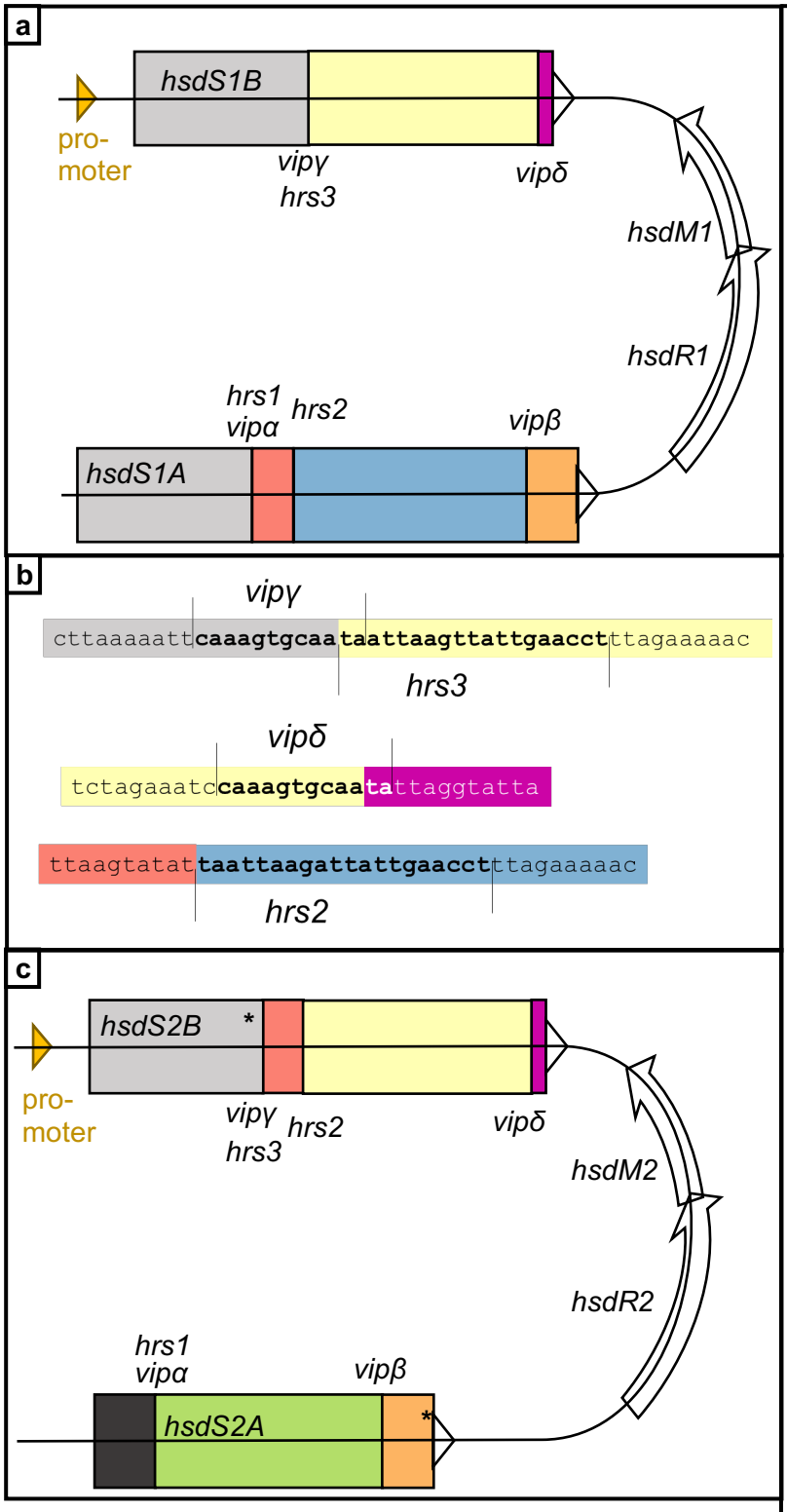
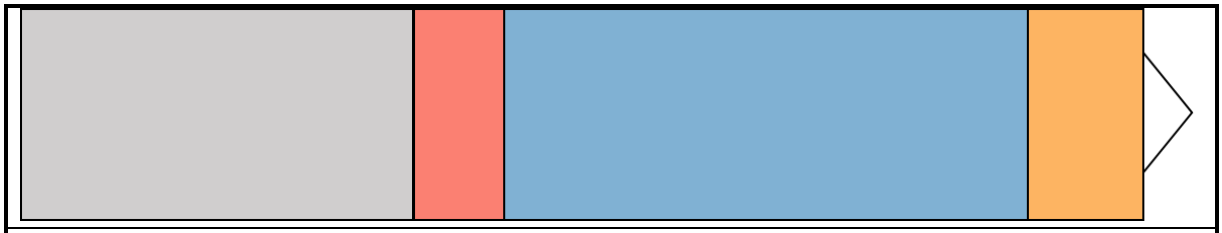


Figure II/4. *hsdS* gene segments.

a Gene segments of *hsd1*, same colours denote sequence identity. Recombination sites are listed at segment junctions. **b** Recombination site sequences and exact positions of gene segment junctions. Colours correlate to gene segments in a. **c** Gene segments of *hsd2*, same colours denote sequence identity to segments in *hsd1*. Stars indicate gene segments with only minor changes (see text). Recombination sites are listed at segment junctions. Note that the black gene segment in *hsdS2A* is never transcribed as it cannot be moved in front of the promoter. For exact gene segment sequences refer to Table II/1.



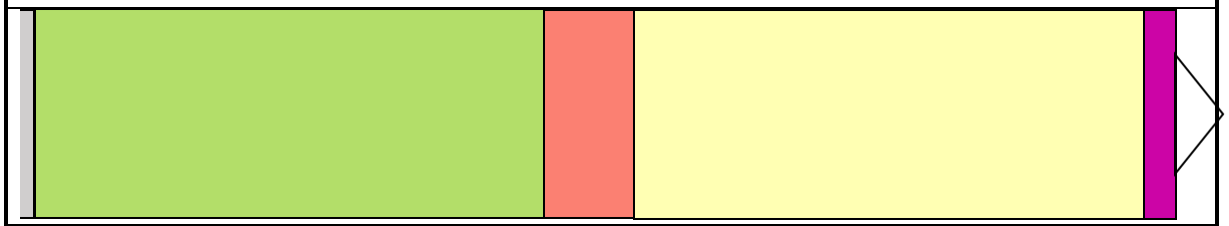
MpuUI gene segment sequences as in *hsd1*:

atggaaatttataaacttggtcagatatt **aaatttagaaaagggcaaatcaaaatata** **atgcaaaatatgtttctcaaaatatcgga** **atttataatttgtactcttcaaaaacaaa** **agatcaaggtatatttggaaaaatcaatt** **catatgactttaatggtgaatatatttta** **attaccactcatggtgcatatgcaggaac** **agttaaatatgtaaatgaaaagttttcca** **caacaagtaattgttttattctaaaagtt** **aatgaaaatattgttaagacaaaatttttt** **aagttatttattattgttacaagaaaaaa** **cattcaatgatatggctataggttctgca** **tatggttatttaaaaaactataacataaa** **tgattttgaagttaatttacctaacttaa** **aaattcaaagt**gcaa	taattaa gattatt gaaccta aagaaga tttattt tttaggc ataaaaa tcttgta agaattg atagtga agaaaat acaaaaa aagattt aagtata t	taattaagattattgaacctttagaaaaacaaataaat gcatttgatgaactgattttgagtgagcaaaaaagtct ccaacattatttaaattatttttttaaataaacttgcat caattaatccttcaattttcaaaaattat**aaacttggt** **gaaatagcaaaaatattaagtggtaaaactccttccac** **tgcaaaaaaggaactatgaaaaaaggaaatacctttttt** **ttggtccgggagatcttgataatatggttccaaaaaga** **tttattacttttaatgaaaaaatgataaaaagatctgg** **caccattttatttttcttctgcagcaacaattgggaaag** **tgggtattttagacaatttatcttgatttaaccaacaa** **ataacttcaatagaggcaaataataactatgttatgga** **taagtttttatttttttcttcttaaaaaaataagttcaa** **aaataaaatttgaaaattcaagcggaacaatatttcct** **acaattaaaaaaaaatattttgaaaattttacactaga** **aattcctaatctaaaaactcaaagt**gcaa	tattaggtat tattgaacca ttgcataaaa aaattaatct tttaaaacaa aagaaaaaat tgcttgaaaa aagatttata tactatcaaa atcacttaat caaggagaaa atcaaagatg agtag

hsd2 gene segment sequences, differences highlighted white

atggaaatttataaacttggtcagatatt **aaatttagaaaagggcaaatcaaaatata** **atgcaaaatatgtttctcaaaatatcgga** **atttataatttgtactcttcaaaaacaaa** **agatcaaggtatatttggaaaaatcaatt** **catatgactttaatggtgaatatatttta** **attaccactcatggtgcatatgcaggaac** **agttaaatatgtaaatgaaaagttttcca** **caacaagtaattgttttattctaaaagtt** **gacgaaaatattgctaagacaaaatttttt** **aagctatttattattgttacaagaaaaaa** **cattcaatgatatggctataggttcagca** **tatggttatttaaaaaactataatataaa** **tgattttgaagttaatttacctaacttaa** **aaactcaaagt**gcaa	As above	As above	tattaggtat tattgaacca ttgcataaaa aaattaatct tttaaaacaa aagaaaaaat tgcttgaaaa aagatctata tactgtcaaa atcacttaat caaggagaaa atcaaagatg agtag

Additional gene segment sequences (MpuT8):



taattaagattattgaacctctagaaaaacaaataaa tgcatttgatgaactgattttgagtgagcaaaaaagt ctccaacattatttaaattatttttttaaataaacttg catcaattaatccttcaattttcaaaaattataaa**ct** **tggtcagatattaaatttagaaaagggcaaatcaaaa** **tataatgcaaaatatgtttctcaaaatatcggaattt** **ataatttgtactcttcaaaaacaagagatcaaggtat** **atttggaaaaatcaattcatatgactttaatggtgaa** **tatattttaattaccactcatggtgcatatgcaggaa** **cagttaaatatgtaaatgaaaagttttccacaacaag** **taattgttttattctaaaagttaatgaaaatattgtt** **aagacaaaatttttaagttatttattattgttacaag** **aaaaaacattcaatgatatggctataggttctgcata** **tggttatttaaaaaactataacataaatgattttgaa** **gttaatttacctaacttaaaaattcaaagtgc**aa	As above	taattaagattattgaacctttagaaaaacaaataaa tgcatttgatgaattgattttgagtgagcaaaaaagt cttcaacattatttgaattatttttttggaaaattct atcaaattgaaccttcattgtttcatgattat**aaact** **cgaaaagattgcaaaaataagaagaggtaagataata** **aattcatttgacctaaaagaaaatcctggagattatc** **ctgtaatttcatcaaatacaaaaaataatggaatttt** **tggttatttaaattcctatatgtatgatggtgagtat** **ataactataagtgcagatggtgcatatgctggaactg** **tgtttttgaataatggaaaatttttctataactaatgt** **gtgtttcattttgttgctaaatgacaaagtaaacctt** **cttacaaaatttctcttttattatttgaaaaagaatg** **aaaatatcatacaaaaaaaatcaatagtaggttcttc** **aagaccatcagttagagaatatactctctcagaaata** **gctatcaaaataccttctctagaaatccaaagt**gcaa	tatt aggt atta atga acac tttc atta tgtt taa

Table II/1. Nucleotide sequences of *hsdS* gene segments.
Nucleotide sequences are shown in a 5'-3' direction. Colours correspond to gene segments of Fig II/4. Differences between *hsd1* and *hsd2* sequences are highlighted in white on a black background. Bold: sequences coding for predicted recognition sites, compare Fig II/5-7.

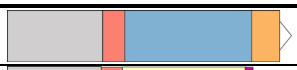
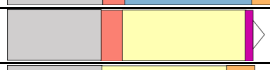
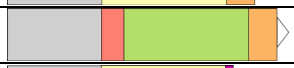
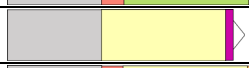
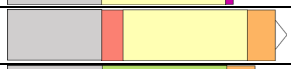
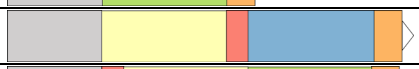
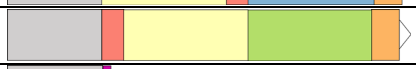
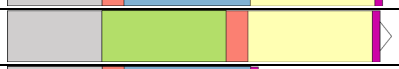
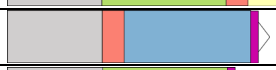
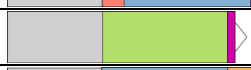
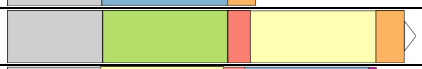
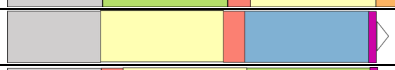
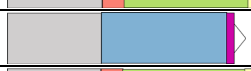
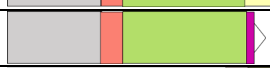
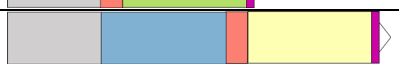
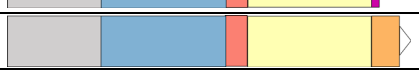
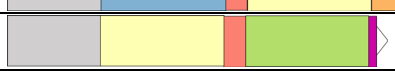
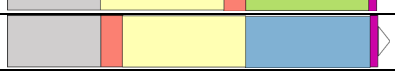
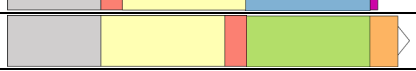
***hsdS* variant**	**Structure**	**Length [nt]**	**Size**
MpuUI		1206	Standard
MpuUII		1110	Standard
MpuUIII		1101	Standard
MpuUIV		1197	Standard
MpuUV		1011	Standard
MpuUVI		1200	Standard
MpuUVIII		1098	Standard
MpuT1		1761	Large
MpuT2		1752	Large
MpuT4		456	Small
MpuT6		546	Small
MpuT7		1671	Large
MpuT8		1662	Large
MpuT9		1116	Standard
MpuT10		1008	Standard
MpuT11		1107	Standard
MpuT14		1752	Large
MpuT15		1671	Large
MpuT16		1662	Large
MpuT17		1761	Large
MpuT18		1662	Large
MpuT19		1017	Standard
MpuT20		1752	Large
MpuT22		1107	Standard
MpuT23		1671	Large
MpuT25		1761	Large
MpuT26		1662	Large
MpuT27		1671	Large
MpuT28		1752	Large
MpuT29		1761	Large

Table II/2. *hsdS* variants produced by the Mpu shufflon.

Colours correspond to gene segments of Fig II/4, for exact nucleotide sequences see Table II/1. Odd-numbered variants can be produced by *hsd1* and even-numbered variants by *hsd2*. MpuU-designated subunits were previously described in Dybvig, Sitaraman and French (1998). Small and Standard sized subunits are most likely to be active *in vivo*.

produced by *hsd2*. Additionally, due to homology of gene segments between loci, we found that several subunit variants can be produced by either *hsd* locus. These include MpuT4/3, MpuT6/5, MpuUV/T12, MpuUII/T13, MpuUVI/T21, and MpuUIII/T24, each referred to by their first designation. When cross-comparing theoretical subunits to previous work, we found that, while MpuT2 and MpuT5 hadn't been sequenced or tested against phage, these *hsdS* confirmations had been found *in vivo* by PCR when analysing a mixed population of *M. pulmonis* (Sitaraman and Dybvig, 1997). This indicates that the diversity generated by the Mpu shufflon indeed reaches further than MpuU-designated subunits found in cultured clones, and demonstrates that our *in silico* constructed MpuT subunits are valid *hsdS* recombinations in *M. pulmonis*. However, this does not verify whether MpuT subunits will produce biologically active HsdS proteins.

To establish which *hsdS* recombinations may be biologically active, we grouped all subunit variants according to their size into 3 categories: small, standard, and large. Small variants range from 456-546 nt and are made up of 2 gene segments, standard variants range from 1008-1206 nt and are made up of 3-4 gene segments, and large variants range from 1662-1761 nt and are made up of 5 segments (Table II/2). This variation in gene lengths indicates that large differences in protein structure may exist for different HsdS variants, and it is likely they will not all be functional. As all previously detailed *hsdS* that are biologically active as a RM specificity subunit are standard sized, MpuT9-11, MpuT19, and MpuT22 (the other five standard sized variants) are the variants most likely to be active. Of these 12 standard sized variants, some vary only slightly in their sequence, with the only difference being the short final gene segment (orange or violet). Should this very short gene segment not be involved in sequence specificity, the number of standard sized variants with unique sequence recognition properties would drop to 6. This would also mean that MpuUIII and MpuUV as well as MpuUII and MpuUVI have the same sequence specificities; from Dybvig *et al*.'s (1998) restriction profiling of HsdS variants it remains uncertain whether these subunits confer different specificities.

In order to make more accurate predictions of the biological activity of different sized variants, we localised target recognition domains (TRDs) to *hsdS* sequences. Previous work predicted the positions for TRDs in *hsd1* by alignment to similar proteins (Sturrock and Dryden, 1997). This allowed us to assign the TRD to the grey (TRD1a), yellow (TRD1b), and blue (TRD1c) gene segments. Due to

gene segment homologies between *hsd1* and *hsd2*, we extrapolated the location of TRD2a (which shares 95% homology with TRD1a) to be within the green gene segment (Fig II/5a and Fig II/6a). As the nucleotide sequence coding for TRD2a is shorter than that coding for TRD1a, we cannot presume that both domains recognise the same DNA sequence.

In order to understand whether all HsdS variants that we generated *in silico* are functional, we modelled various subunits as 3-dimensional HsdS proteins using Phyre2 (Kelly *et al.*, 2015). These homology-directed protein models are largely based on the solved crystal structure of a Type I RM specificity subunit found in *Methanococcus jannaschii* (Kim *et al.*, 2005). Standard-sized HsdS subunits such as MpuUI are predicted to fold into a 46 kDa protein with rotational symmetry (Fig II/5 left). The 5'-most grey gene segment coding for TRD1a gets translated as a globular DNA-binding domain. In the case of MpuUI, a helical spacer domain derived from the red gene segment forms a bridge between this first and a second globular DNA binding domain, which is derived from the TRD1c-encoding blue segment. The final orange segment codes for a second helical spacer domain, which spans the gap between DNA-binding domains parallel to the red spacer domain. Both globular domains have a small groove on the face furthest from linker domains, which is perhaps the DNA-binding section of the protein. In HsdS variants that lack a gene segment corresponding to the helical linker between globular domains, e.g. MpuT11 (Fig II/5 right), a portion of the second globular domain is predicted to adopt a helical shape to become a spacer domain. This is highly likely to change DNA specificity, as it requires a change in shape of globular domain. However, as homology-directed models can have severe limitations in accuracy of structures more detailed than overall protein shape, we suggest an alternative model (Fig II/5c left) in which the aforementioned portion of the second globular domain (blue) always forms a short helical spacer. In this alternative model, presence of an additional spacer domain would result in a protein with increased spacing between globular DNA-binding domains whilst leaving the DNA interaction sites unchanged, which will nevertheless change the DNA recognition sequences because spacing between recognized sequences is altered. In contrast, HsdS variants that lack a TRD-coding gene segment (e.g. MpuT6; Fig II/7 left), form a truncated protein. In other Type I RM systems, truncated HsdS proteins with only one TRD can dimerise to form a functional specificity subunit

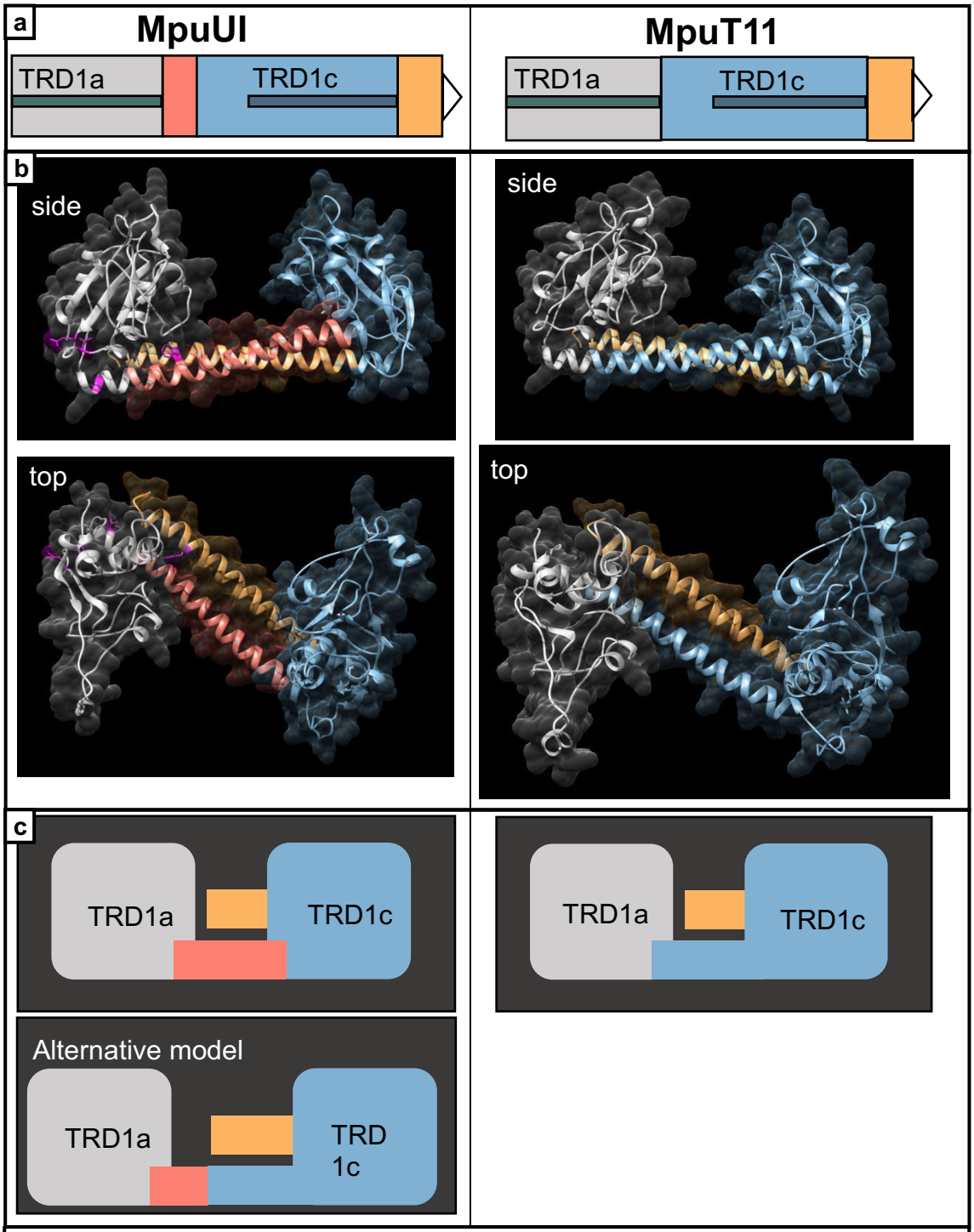


Figure II/5. Structure of MpuUI and MpuT11 genes and their proteins.

a Gene structures with positions of Target Recognition Domain (TRD)-coding regions superimposed. **b** Homology-directed protein models. Colours correspond to gene segments in (a). MpuUI model shows positions of amino acids which are substituted in *hsd2* gene segments in pink. Note that the orange segment contains two substitutions, one of which is omitted in the model. **c** Simplified models of HsdS proteins based on models in (b). For description of alternative models, see text.

(Meister *et al.*, 1993). Acting in this way, MpuT4 or MpuT6 could become functional subunits with an entirely new sequence specificity by dimerising with

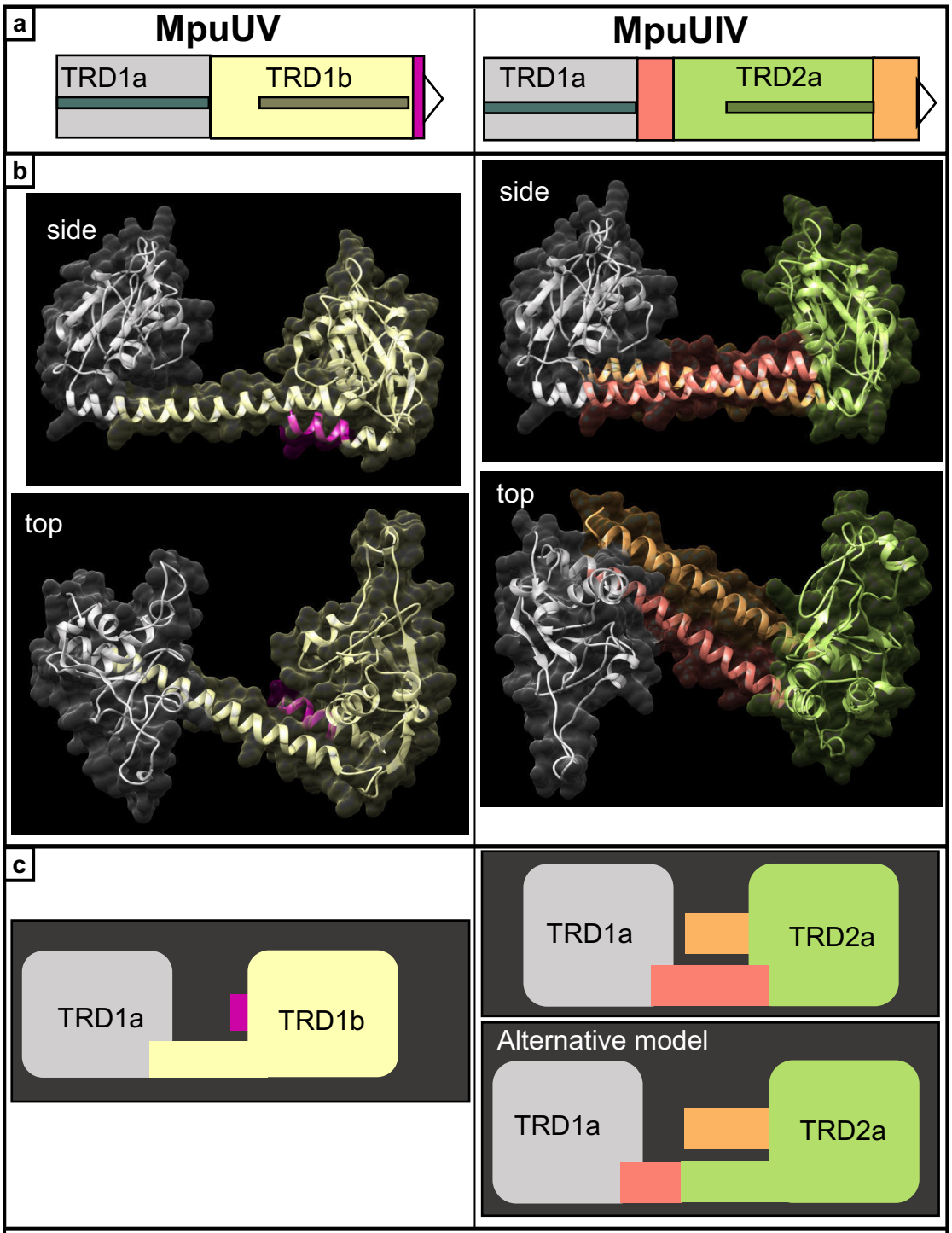


Figure II/6. Structure of MpuUV and MpuUIV genes and their proteins.
a Gene structures with positions of Target Recognition Domain (TRD)-coding regions superimposed. **b** Homology-directed protein models. Colours correspond to gene segments in (a). **c** Simplified models of HsdS proteins based on models in (b). For description of alternative models, see text.

themselves or each other. In HsdS variants with an additional TRD-coding segment (e.g. MpuT2; Fig II/7 right), the first linker domain (red) is predicted to fold up on

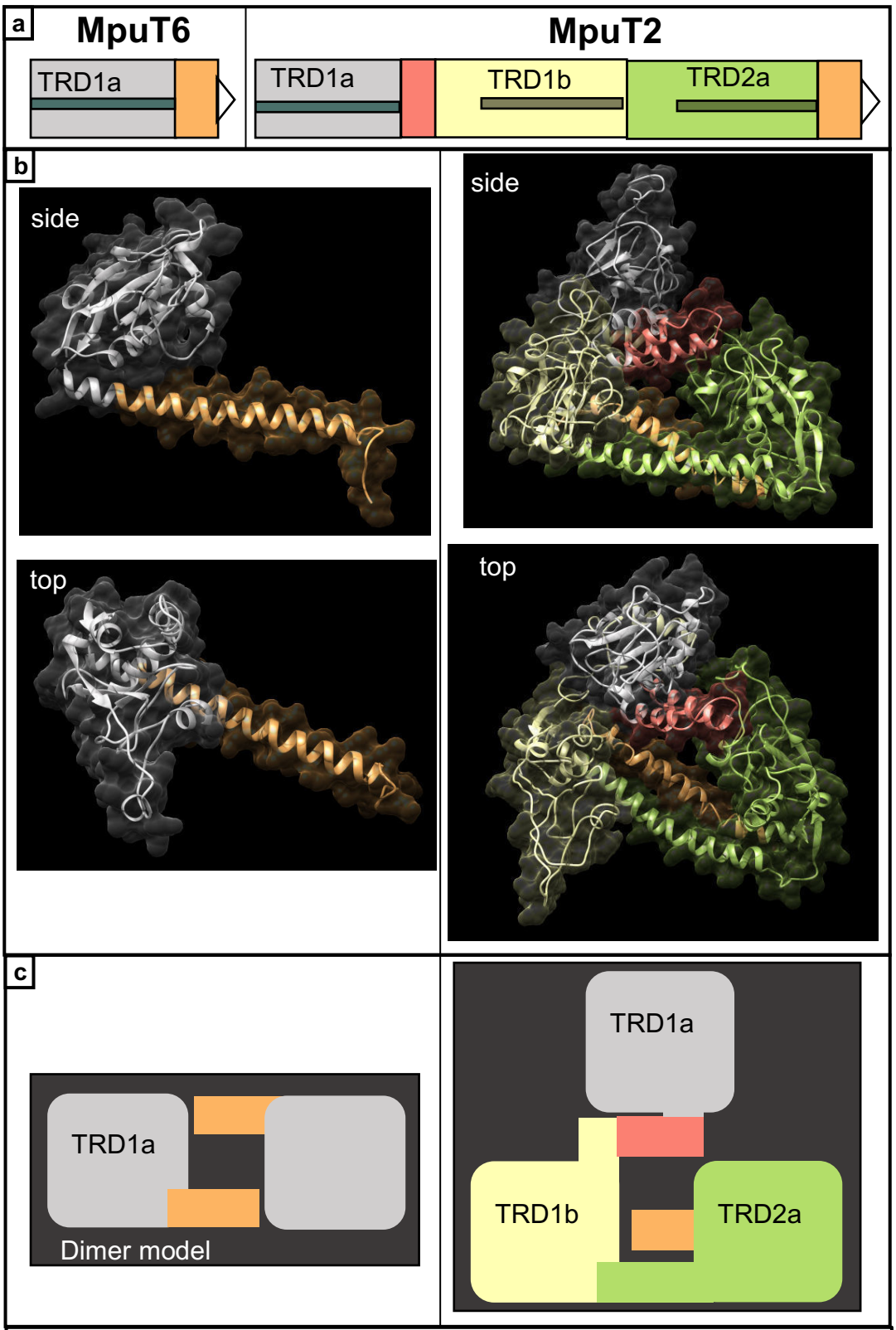


Figure II/7. Small and large sized HsdS proteins.

Left: MpuT6, Right: MpuT11 **a** Gene structures with positions of Target Recognition Dmain (TRD)-coding regions superimposed. **b** Homology-directed protein models. Colours correspond to gene segments in (a). **c** Simplified models of HsdS proteins based on models in (b). Left: MpuT6 dimer model.

itself and adopt a U-shaped helical structure, which places the second globular domain directly underneath the first. This results in a protein in which the second TRD-containing segment (yellow) adopts the role of the first globular domain, with the remaining TRD-containing segment folding into a spacer domain as well as second globular domain. The surplus globular domain (grey) protrudes from the top of the otherwise standard HsdS structure and blocks access to the other DNA-binding domains. Therefore, if the additional globular domain is not removed in post-translational processing, HsdS variants containing three TRDs are likely to be non-functional.

Homology-based protein modelling also has its limits, however. Phyre 2 is likely not to change the structure of the coiled coil spacer domains, even when their sequence changes. Therefore, these models are likely to under-estimate the amount of diversity that can be generated by the Mpu RM shufflon, in particular within the spacer region between TRDs.

Finally, we mapped amino acid changes between grey and orange *hsd1* and *hsd2* gene segments to our model of MpuUI to determine whether our earlier assumption that these mutations do not change gene segment properties can be upheld (Fig II/5b). Firstly, the orange segment is unlikely to show altered properties as it forms a spacer domain; their properties are largely resistant to amino acid changes (Cowan, Gann and Murray, 1989). Secondly, only two of the three amino acid changes in the TRD-containing grey gene segment are at the protein surface, but they are not on the protein's DNA-facing side. Rather, they can be found in the interface of TRD and spacer domains, a region typically associated with binding of HsdM (Zinkevich *et al.*, 1992). This indicates that the differences in *hsd1* and *hsd2* gene segments may be for optimisation of association with the respective HsdM1 or HsdM2 methyltransferase. Therefore, as HsdS subunits from either *hsd* locus can associate with either methyltransferase and restriction endonuclease, we can disregard the minor differences between grey and orange *hsd1* and *hsd2* gene segments as it is highly likely that these do not produce functionally different proteins.

In summary, we found that the Mpu shufflon *hsd* loci have the capacity to produce 30 different *hsdS* genes that can be grouped into small, standard, and large sized subunits. Standard sized variants include all previously described MpuU HsdS proteins and contain two TRDs. Small sized variants encode one TRD and may dimerise to form a functional HsdS complex, while large sized variants contain

three TRDs and are unlikely to form functional specificity subunits. Overall, the Mpu shufflon can encode 14 different standard and small sized subunits, all of which are likely to be functional. With our database of nucleotide sequences for all *hsdS* variants of the Mpu shufflon, it will be easier to, in future, express and analyse these proteins *in vitro* to answer structural questions, such as which gene fragments are essential for a change in specificity, and whether HsdS with three TRDs can be functional.

In *M. pulmonis* and other shufflons, recombination occurs in a very rapid timeframe, so that clones with different genotypes can be found within a bacterial population (Dybvig, Sitaraman and French, 1998; Tettelin *et al.*, 2001). This demonstrates that shufflons are a means of generating population-level RM diversity, such a population with multiple RM specificities will be resistant to methylated escape phage sweeping through the entire population. A different bacterial diversity-generating immune mechanism, the adaptive response CRISPR-Cas, functions by integrating short phage DNA sequences into its own genome as spacers, and later targeting phage that carry the same sequence in their genome. In comparison to RM shufflons, CRISPR has the capacity to generate far more diversity by integrating a large range of different spacers. An experiment tracking emerging CRISPR escape phage showed not only that this diversity can rapidly drive phage extinct, but also indicated that there may be a threshold level past which additional CRISPR diversity is redundant (van Houte *et al.*, 2016). If a shufflon can produce enough diversity to pass this threshold, it may limit phage epidemics and the emergence of escape phage just as effectively as CRISPR-Cas. Another key difference between these two immune responses is the mechanism of phage escape. To escape RM, a phage needs to become methylated. This can occur by chance, or can be enabled by phage-encoded proteins (Krüger and Bickle, 1983). Alternatively, some phage encode anti-restriction mechanisms to directly block aspects of RM (Labrie, Samson and Moineau, 2010). To escape CRISPR, phage need to acquire point mutations in the region of their genome targeted by the bacteria's spacer, which can be costly to the phage if the spacer targets an essential gene. However, a phage can become resistant to multiple CRISPR spacers by acquiring several point mutations, while it can only be resistant to one RM diversity at a time. Therefore, if a RM shufflon can produce enough diversity, it may be a more effective means of limiting phage infections in

a bacterial population than CRISPR-Cas. However, costs of RM shufflons to bacteria need to be considered too. As RM systems cleave unmethylated DNA, a switch in DNA sequence specificity may be toxic as new recognition sequences will be unmethylated on host DNA. While several restriction alleviation mechanisms exist (Murray, 2000), none are known in *M. pulmonis*. These speculative costs of switching specificity in shufflons may make CRISPR the more feasible immune system for population-level diversity and would help explain why RM shufflons are not more common throughout bacteria.

In order to approach these evolutionary questions, we aim to set up an experimental system to study phage-bacteria coevolution in presence of RM shufflons as well as interactions between RM and CRIPSR in the next chapter while also addressing RM costs. Encoding the Mpu shufflon on artificial gene constructs with fixed *hsdS* variants will allow us, in future, to generate a bacterial population with various RM specificities in a very controlled way.

Methods

REBASE analysis

For an analysis of the presence of multiple *hsdS* genes in Type I RM systems, we downloaded all Type I RM genes from REBASE as of 22/9/16 (Roberts *et al.*, 2015). Genes on REBASE are listed under their systematic names; *hsdR* with a unique species/strain/gene identifier, e.g. EcoKI, *hsdM* with the prefix “M.”, e.g. M.EcoKI, and *hsdS* with the prefix “S.”, e.g. S.EcoKI. If more than one *hsdS* gene is present for an RM system, they are appended with the prefixes “S1.”, “S2.”, etc. We searched for genes starting with “M.” or “M1.” to return the number of methylases listed as an indication of number of unique RM systems. To determine the number of RM systems with >1 *hsdS* subunit, we searched for genes starting with “S2.” A search for genes beginning with “S3.” gave us those with >2 *hsdS* subunits, and so on.

hsdS sequences

Nucleotide sequences for *hsdS* genes MpuUI-MpuUVIII were found on Genbank (Clark *et al.*, 2016) under accession numbers AF076984 (MpuUI), AF076985 (MpuUII), AF076986 (MpuUIII), AF076987 (MpuUIV), AF076988 (MpuUV), AF076989 (MpuUVI), and AF076990 (MpuUVIII). These sequenced were annotated for *vip* and *hrs* sites, from which gene segments could be deduced. We

defined segments to split at the last two nucleotides of *vip* sites (5'-CAAAGTGCAA-TA-3'), and immediately upstream of *hrs* sites. This ensures that gene segments can be seamlessly reshuffled *in silico* without changing the gene's reading frame, even at gene segment junctions where *vip* and *hrs* sites overlap.

Theoretical HsdS subunits were built by generating a database with *hsdS* gene segment sequences and the *hsd* loci architecture. Systematically, we simulated inversions between matching recombination sites. To verify *in silico* recombination, MpuU variants generated *in silico* were compared to reported subunits using BLAST (Altschul *et al.*, 1990), and their sequence uniformity confirmed (100% identity).

TRD mapping and protein models

TRDs in MpuU proteins predicted by Sturrock and Dryden (1997) were mapped onto translations of *in silico* generated Mpu variants, obtained by ExPASy's translate tool (Artimo *et al.*, 2012) and the nucleotide sequence coding for each TRD tracked back onto gene segments. The alignment between TRD1a and TRD2a was carried out with BLAST (Altschul *et al.*, 1990).

We produced protein fold models of HsdS variants with Phyre2 (Kelly *et al.*, 2015) in normal modelling mode; to ensure that the entire amino acid chain was used to predict the model intensive modelling mode was used for MpuT2 and MpuT4. We coloured protein models according to gene segments and mapped point mutations using UCSF Chimera (Pettersen *et al.*, 2004).

Chapter III: Restriction-Modification in *Pseudomonas aeruginosa*

Abstract

Restriction-Modification (RM) is one of the most widespread bacterial immune mechanisms, but in many cases it is unclear what the costs and benefits of this system are. This is especially true for Type I RM shufflons, which rapidly recombine their specificity subunit to generate a new sequence specificity. We generated synthetic gene constructs coding for the Mpu shufflon found in *Mycoplasma pulmonis*, and this shufflon was expressed in a *Pseudomonas aeruginosa* PA14 strain lacking CRISPR-Cas to establish a model system for RM study. In unmethylated hosts, introduction of this new RM system is toxic due to self-DNA cleavage. By initially transforming with an RM construct lacking *hsdR* (the restriction endonuclease) and later introducing the full RM system, this toxicity can be circumvented. Even though the Mpu RM proteins are active in *P. aeruginosa*, they only confer marginal levels of resistance against *Pseudomonas* phages JBD5 and Φ1214. This gives an indication how introduction of a novel RM system may be associated with greater costs than benefits.

Introduction

As one of the most widespread bacterial immune systems, Restriction-Modification (RM) is crucial for interactions of many bacteria with bacteriophage (phage) and other mobile genetic elements. However, RM not only provides a benefit through cleaving phage, these systems also methylate the genomes of their bacterial hosts. In some cases, this is a means of gene regulation through methylome maintenance and the main function of the RM system (Ishikawa, Fukuda and Kobayashi, 2010), but in other cases methylation patterns could provide little or no benefit, or even be costly to a bacterium. Unexpected methylation patterns can cause pleiotropic costs which are more difficult to quantify. Furthermore, autoimmunity through self-cleavage can attribute an obvious cost to carrying a RM system. This trade-off between costs and benefits becomes especially pertinent when bacteria acquire a new RM system, which is when autoimmunity or altered methylation patterns new to that bacterium are most likely to have an effect.

The Mpu shufflon found in *Mycoplasma pulmonis* (Dybvig, Sitaraman and French, 1998) encodes a Type I Restriction-Modification (RM) system that rapidly recombines its specificity subunit, *hsdS*, as described in the previous chapter. In such a shufflon, hosts will constantly be exposed to novel sequence specificities and may therefore be subject to the same costs associated with introducing an entirely novel RM system.

As an immune system, RM needs to distinguish between self and non-self DNA to only cleave invasive DNA. RM achieves this by methylating the bacterial chromosome. In instances where this fails, the bacteria's own genome can be cleaved through autoimmunity (Pleška *et al.*, 2016), therefore ensuring genome methylation is crucial. Vertical maintenance of methylation within a bacterium and its daughter cells is fairly straightforward. Daughter cells' DNA will be methylated on one DNA strand only, such hemimethylation can induce the methylase to methylate the opposing DNA strand in the same pattern (Vovis and Zinder, 1975). The issue of self-restriction becomes more difficult when RM systems get transferred horizontally, which has occurred extensively throughout evolution (Nobusato, Uchiyama and Kobayashi, 2000; Kobayashi, 2001). How, if it is in a new host, can a RM system distinguish between self and non-self DNA?

There are several mechanisms of restriction alleviation (RA) for RM systems. Generally, RA causes a delay of restriction activity, leaving enough time for the methylase to methylate the bacterial genome. In some Type I RM systems, the protein complex itself causes this necessary lag by an unknown post-translational mechanism (Kulik and Bickle, 1996). In *E. coli*, *hsdR* has its own promoter, and in this way its activity can be regulated separately from *hsdM* and *hsdS*. However, this is not the case in *Mycoplasma pulmonis*, and currently there is no described mechanism of RA for its shufflon. Therefore, we cannot predict whether changing specificity in this shufflon will be associated with large autoimmune costs that require RA.

To simulate the costs of autoimmunity that might be associated with the introduction of a new RM specificity, and also to investigate benefits in the form of phage resistance, we attempt to transfer the Mpu shufflon to *Pseudomonas aeruginosa* in this chapter. Additionally, this groundwork will establish a model system for further study of how bacterial populations with diverse RM specificities co-evolve with their phages, and allow investigation of the interplay between RM and CRISPR-Cas systems. To do this, we will design gene constructs to express the

Mycoplasma pulmonis Mpu shufflon, with fixed specificities, in *P. aeruginosa* PA14.

Results & Discussion

With the great evolutionary distance between *P. aeruginosa* and *M. pulmonis*, it is unlikely that PA14 has ever encountered the Mpu shufflon or a similar RM system. Therefore, when transferring the shufflon into a *P. aeruginosa* host, we will likely see the full costs associated with a novel RM system. To investigate the extent of these, we generated a gene construct *in silico* that codes for one variant of this shufflon (Fig III/1a).

This synthetic "SRM construct" was designed to code for the proteins HsdS (in the MpuUI conformation), HsdM1, and HsdR1 exactly as they are produced in *M. pulmonis*, but the genes themselves were separated and their codon usage optimised for *P. aeruginosa* PA14. Restriction sites as well as predicted ribosome binding sites (RBS) within the genes were removed and an artificial T7 RBS, the same as found on expression vector pHERD-30T, introduced in the non-coding region preceding each gene. The *hsdS* gene was flanked by NcoI and KpnI restriction sites to allow modular exchange with constructs coding for alternative variants of HsdS. BamHI was introduced as a restriction site downstream of *hsdM* to allow insertion of the entire SRM construct into pHERD-30T (pHERD30T_SRM; Fig III/1b). In this way, the Mpu shufflon can be expressed from an arabinose-inducible promoter (Qiu *et al.*, 2008) in *P. aeruginosa* PA14. Additional "S constructs", coding for all other standard sized HsdS subunits (MpuUII-MpuUVI, MpuUVIII, MpuT9-MpuT11, MpuT19, MpuT22) were created *in silico* as above and flanked by NcoI and KpnI restriction sites. In future, synthesising these S constructs on separate plasmids will allow for easy directed expression of specificity subunits when cloning a new *hsdS* gene into SRM-30T using the flanking restriction sites. The main plasmids used throughout this study, pHERD30T_SRM and pHERD30T_SM, are summarised in Figure III/1c. We transformed *P. aeruginosa* PA14 *lacZ::csy3* with these plasmids to generate RM transformant strains Csy3::Lacz pHERD30T_SRM and Csy3::LacZ pHERD30T_SM respectively.

To understand if the RM system was functionally expressed, we tested the *in vitro* restriction abilities of Mpu shufflon proteins in lysate. Accordingly, we isolated

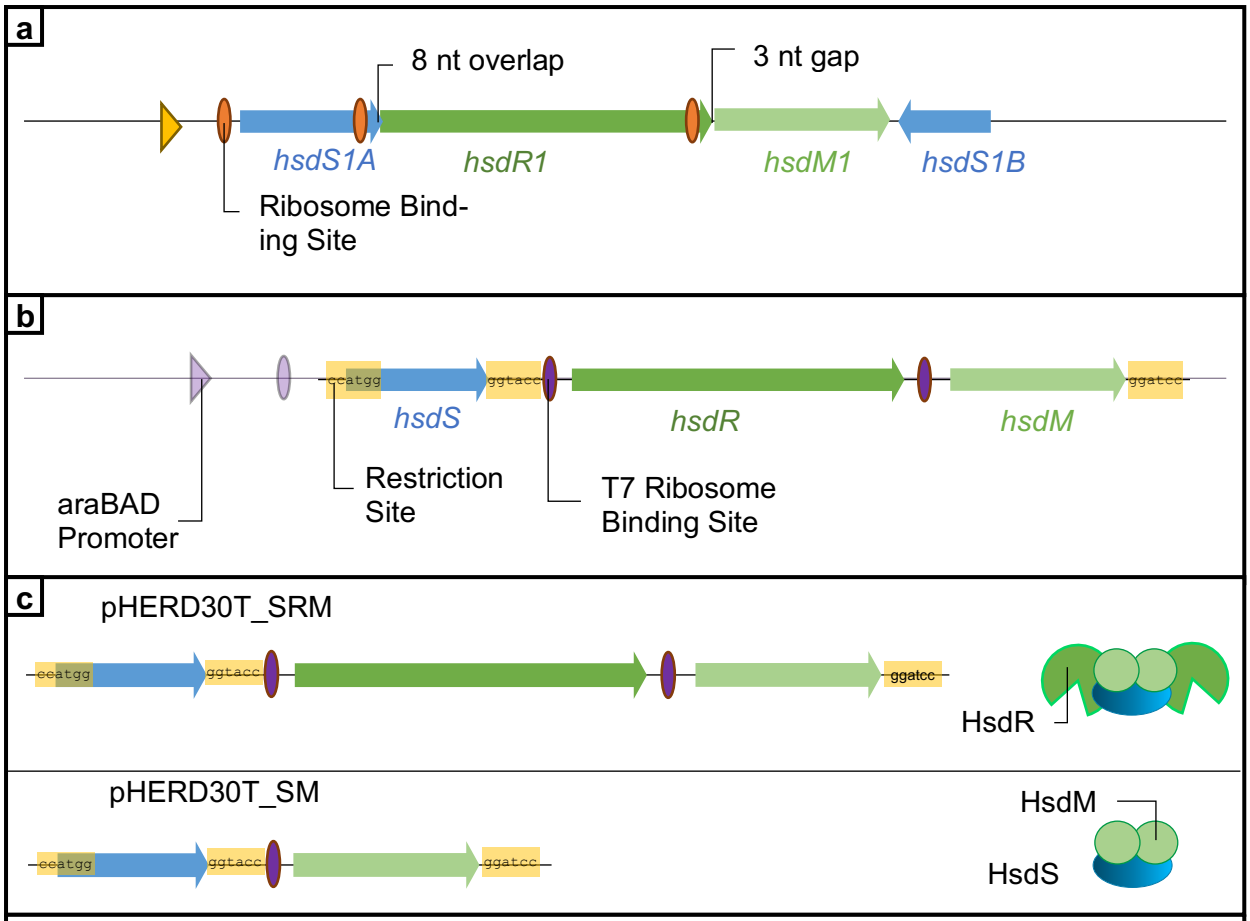


Figure III/1. Synthetic gene constructs coding for the *M. pulmonis* Mpu shufflon. **a** Gene locus *hsd1* as found on *M. pulmonis* in inverted orientation (Fig II/3b; Genbank acc. number L25415). **b** Genes were seperated and codon-optimised for *P. aeruginosa* PA14. Synthetic T7 ribosome binding sites and restriction sites were introduced between genes. Semi-transparent elements indicate the promoter and RBS found on the expression vector pHERD-30T. **c** Synthetic gene constructs (left, expression vector not shown) and their protein products (right). pHERD30T_SRM codes for the complete RM system with specificity subunit MpuUI, pHERD30T codes for HsdS (MpuUI) and HsdM.

genomic DNA (gDNA) from PA14 WT as well as *Pseudomonas* phage Φ1214 and incubated it together with protein lysate obtained from Csy3::LacZ pHERD30T_SRM where RM expression had been induced by arabinose. Results show that SRM lysate induces cleavage of both PA14 WT and Φ1214 gDNA overnight (Fig III/2). The empty vector control as well as Csy3::LacZ pHERD30T_SM lysate do not induce DNA cleavage in the same timeframe. More efficient cleavage might have been achieved by using more concentrated reaction buffer, or by optimising lysate extraction or restriction reactions. These data demonstrate that the Mpu shufflon can be functionally expressed in *P. aeruginosa*

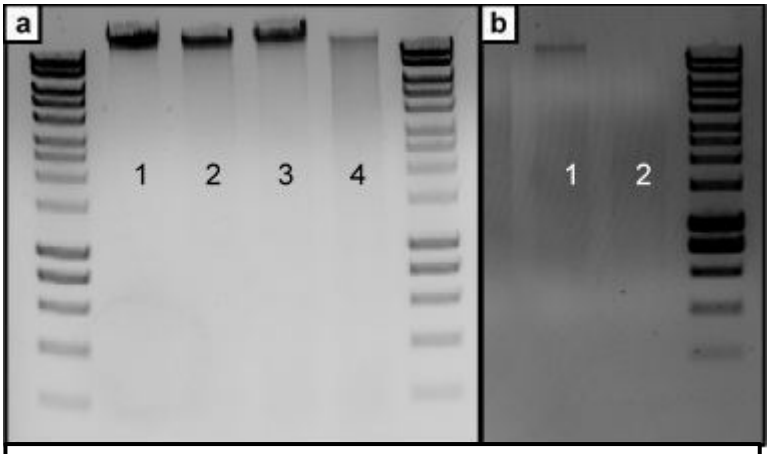


Figure III/2. pHERD30T_SRM lysate can cleave genomic DNA.

Genomic DNA cleavage induced by protein lysate after overnight incubation at 37°C. **a** Cleavage of 1 µg PA14 WT gDNA with lysate isolated from 1-no-lysate control, 2-Csy3::LacZ pHERD30T, 3-Csy3::LacZ pHERD30T_SM, 4-Csy3::LacZ pHERD30T_SRM. Similar results were obtained in three experiments. **b** Cleavage of 300 ng Φ1214 gDNA with lysate isolated from 1-Csy3::LacZ pHERD30T, 2-Csy3::LacZ pHERD30T_SRM.

from an expression plasmid, and how the RM proteins show activity in this bacterial host. As expected, HsdR is an essential component for DNA cleavage. Furthermore, MpuUI has the capacity to restrict Φ1214. Typical Type I RM DNA recognition sites contain three nucleotides, a stretch of unspecific nucleotides, and another four specific nucleotides that are required to match. Therefore, in a random DNA sequence, we would expect one recognition site every 16.4 kb. As two recognition sites are required for cleavage, we expect phages with a genome of over 33 kb to be targeted by Type I RM without knowing its recognition sequence. Φ1214 has one of the smallest genome sizes of phages used throughout this thesis (~37 kb, Table S2), therefore we can tentatively predict all *Pseudomonas* phages to be targeted by MpuUI. Additionally, as these results show that the Mpu RM system has the capacity to cleave unmethylated PA14 DNA, they suggest that an induction of RM expression may be toxic due to autoimmunity.

To examine this hypothetical toxicity of RM we first compared the efficiency of transformation (EOT) of plasmids encoding the RM system to the EOT of the empty vector. Strikingly, EOT of pHERD30T_SRM, but not of the empty vector, is drastically decreased at 1% (w/v) arabinose, compared to transformations without arabinose (Fig III/3). These data suggest that RM expression is toxic, presumably because the host DNA is cleaved by the novel RM system. We also observed that Csy3::LacZ pHERD30T_SRM strains grew very poorly on plates

(not shown), further suggesting that RM expression resulted in cytotoxicity. To examine this in more detail, we measured Csy3::LacZ pHERD30T_SRM growth in the presence of arabinose over 16 hours by measurement of OD600. At both moderate (0.1% (w/v)) and high (1% (w/v)) arabinose concentrations, growth of bacteria transformed with pHERD30T_SRM was drastically reduced compared to those with empty vectors (Fig III/4). While growth for Csy3::LacZ pHERD30T_SM at 1% (w/v) arabinose is reduced compared to empty vector control bacteria, it is considerably higher compared to pHERD30T_SRM transformants and might merely be due to the cost of protein overproduction. Importantly however, growth at 0.1% (w/v) arabinose for Csy3::LacZ pHERD30T_SRM is extremely limited, whereas it appears normal for pHERD30T_SM transformants. This demonstrates the toxicity of the restriction subunit HsdR. Together with the EOT results, the above data indicate that the transformation of this new RM system is toxic to Csy3::LacZ.

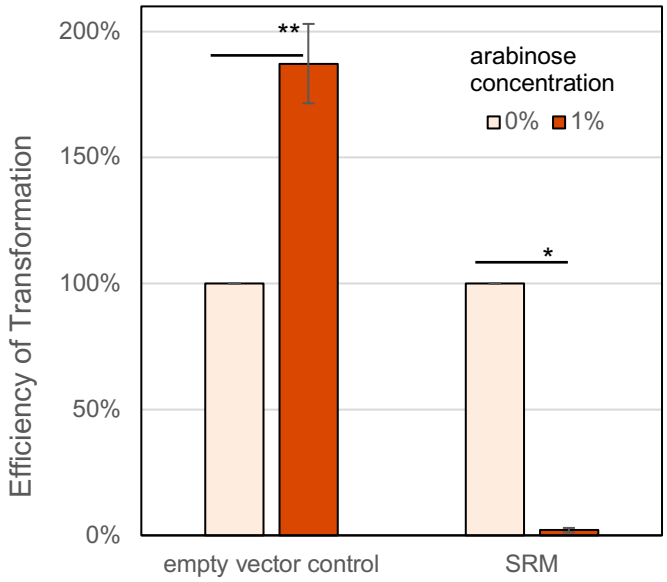


Figure III/3. pHERD30T_SRM reduces EOT in the presence of arabinose.
Efficiency of Transformation (EOT) of Csy3::LacZ with pHERD30T_SRM or empty vector control was calculated by counting colonies on transformation plates, with colonies on 0% (w/v) arabinose plates being set as 100%. Shown are means of 4-5 experiments with their standard error. Significance tested by Wilcoxon-ranked sums for each treatment. *$p<0.05$ ($p\approx0.021$); **$p<0.01$ ($p\approx0.0075$)

To test whether Csy3::LacZ pHERD30T_SRM bacteria could adapt to their constructs, thus circumventing RM toxicity, we gradually

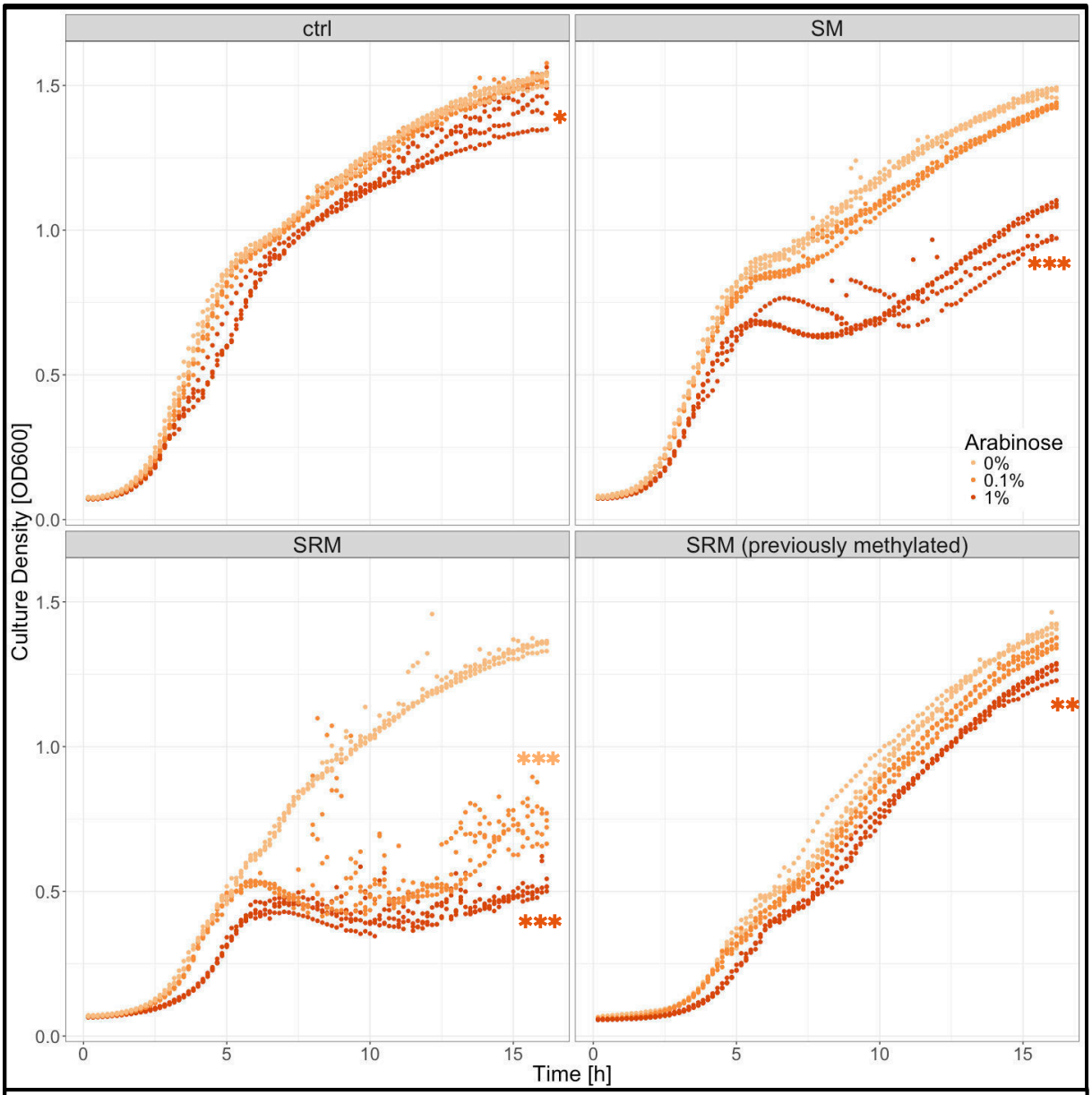


Figure III/4. pHERD30T_SRM causes reduced growth in unmethylated strains. Culture density measurements as a proxy of bacterial growth at different arabinose concentrations over a course of 16 hours. All data points are shown. Ctrl: Csy3::LacZ pHERD30T, *p≈0.0034; SM: Csy3::LacZ pHERD30T_SM, ***p≈5x10^{-5}; SRM: Csy3::LacZ pHERD30T_SRM (none previously induced), ***p<10^{-7}; SRM (previously methylated): Csy3::LacZ where the genome was methylated before transformation with pHERD30T_SRM, **p≈0.0039. N=4, similar results were obtained for each strain in 2-3 additional experiments. P values refer to differences to 0% arabinose treatment, see Methods.

transferred these transformants to incrementally higher arabinose concentration in three replicates (Fig III/5a). When plating these Csy3::LacZ pHERD30T_SRM onto 1% (w/v) arabinose, we observed a considerably lower number of colonies compared to empty vector control in the early stages of this adaptation. In later

stages of adaptation, the number of colonies of Csy3::LacZ pHERD30T_SRM on 1% (w/v) arabinose rose to nearly 95% of colonies of empty vector controls transferred in the same way (Fig III/5b). This shows that Csy3::LacZ can become accustomed to SRM toxicity if it is incrementally adapted to high arabinose concentrations over four days. Additionally, to evaluate whether these adapted bacteria would show the same growth restrictions as un-adapted Csy3::LacZ pHERD30T_SRM, we measured their growth over 16 hours as before. Adapted Csy3::LacZ pHERD30T_SRM of all three replicates show growth curves indistinguishable from empty vector control bacteria with and without arabinose (Fig III/6). Together, these data suggest that by slowly increasing pHERD30T_SRM expression, RM toxicity can be circumvented.

There are two plausible explanations for this alleviation of toxicity; either some form of RA occurred and Csy3::LacZ transformants successfully became methylated, or the construct was so toxic to bacteria that RM mutants were selected for. A loss of HsdS or HsdR function, or a complete loss of HsdM would result in a RM system unable to cleave DNA, and therefore alleviate toxicity. In order to resolve whether this alleviation of SRM toxicity was due to a loss of RM function, we tested infectivity of a range of phages on adapted Csy3::LacZ pHERD30T_SRM. In the three replicates the efficiency of plaquing (EOP) of phages used did not significantly differ from their EOP on empty vector control bacteria (Fig III/5c). This indicates that adapted pHERD30T_SRM does not confer immune functions to PA14. Additionally, we determined EOTs for adapted pHERD30T_SRM constructs by isolating plasmid DNA from bacterial cultures after the four-day transfers. Two out of three replicates show a significantly higher EOT than ancestral pHERD30T_SRM at 1% (w/v) arabinose and are on par with the empty vector control (Fig III/5d). These data show that after undergoing adaptation, the SRM construct has lost its properties. While the mean EOT for the final replicate (A) is also higher than EOT of ancestral pHERD30T_SRM, this difference is non-significant due to high variation. In summary, while construct toxicity is lost when adapting Csy3::LacZ pHERD30T_SRM to higher arabinose concentrations, this is due to loss of RM function rather than host DNA methylation. In order to determine which RM component mutates to make the immune system non-functional, we could in future studies isolate plasmid DNA from these strains for re-sequencing.

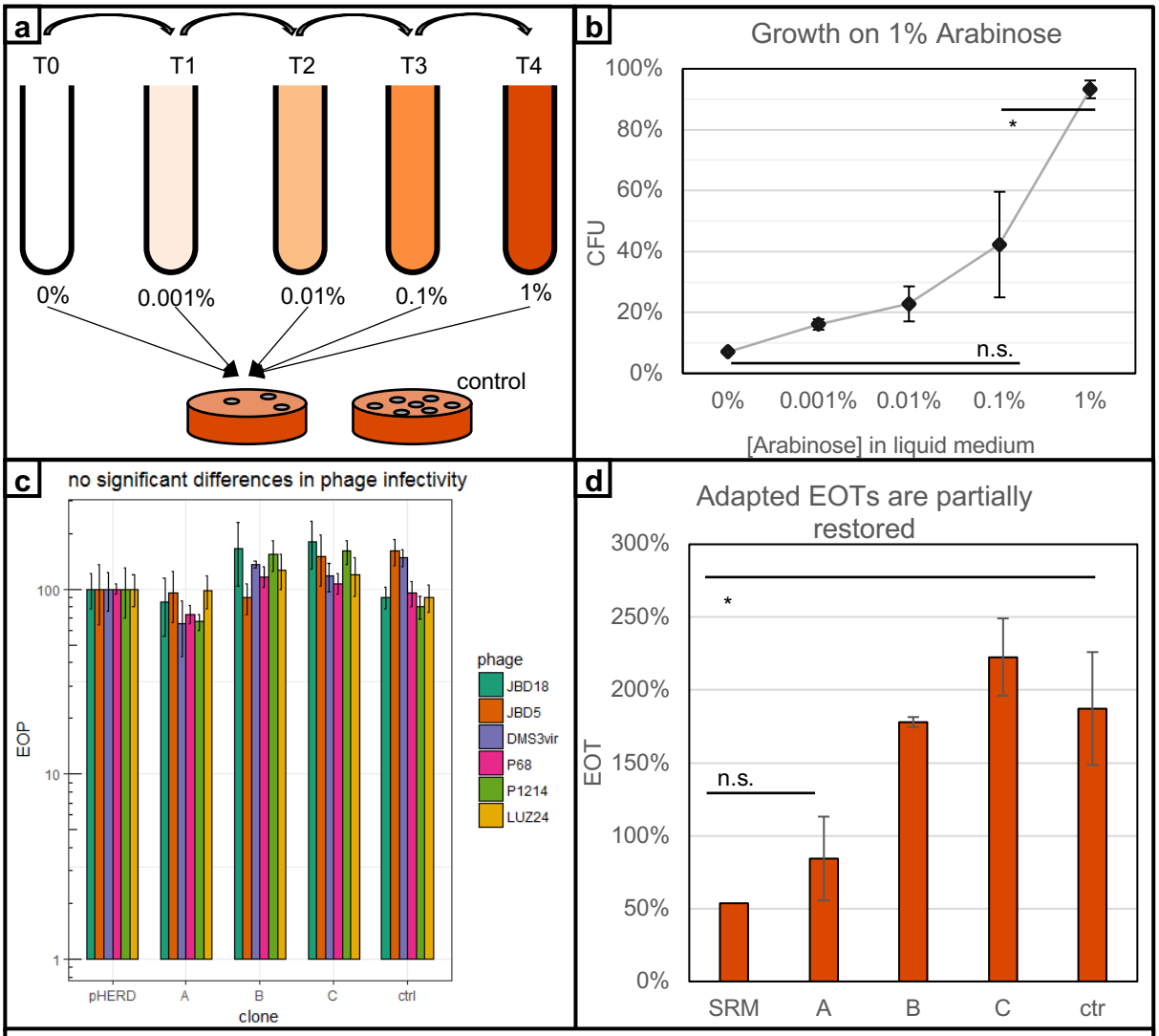


Figure III/5. Adaptation to high arabinose concentrations restores Csy3::LacZ pHERD30T_SRM viability but causes loss of RM function.

a Csy3::LacZ pHERD30T_SRM (not previously induced) was transferred in three replicates with a daily increasing arabinose concentration and plated onto 1% (w/v) arabinose plates. **b** Colony forming units (CFU) relative to the control throughout the experiment. **c** Efficiency of Plaquing (EOP) as a measure of phage infectivity compared to infection of an empty vector control. pHERD - control strain, not from experiment. A, B, C - colonies from the three replicates at T4. ctrl - colony from control treatment at T4. N = 3. No significant differences to pHERD were detected. **d** Plasmid DNA extracted from cultures A, B, C, ctrl as described in (c) was used to transform Csy3::LacZ. Efficiency of Transformation (EOT) at 1% (w/v) arabinose was calculated relative to that of the respective plasmid at 0% (w/v) arabinose. SRM – ancestral pHERD30T_SRM isolated from *E. coli*. N = 2-5. **b, c, d** Shown are means with their respective standard errors. Significance was tested by analysis of variances and post-hoc Tukey's multiple comparison of means. * statistically significant ((b) $p \approx 0.034$, (d) top to bottom: $p \approx 0.016$, 0.0074, 0.027, n.s.: not significant

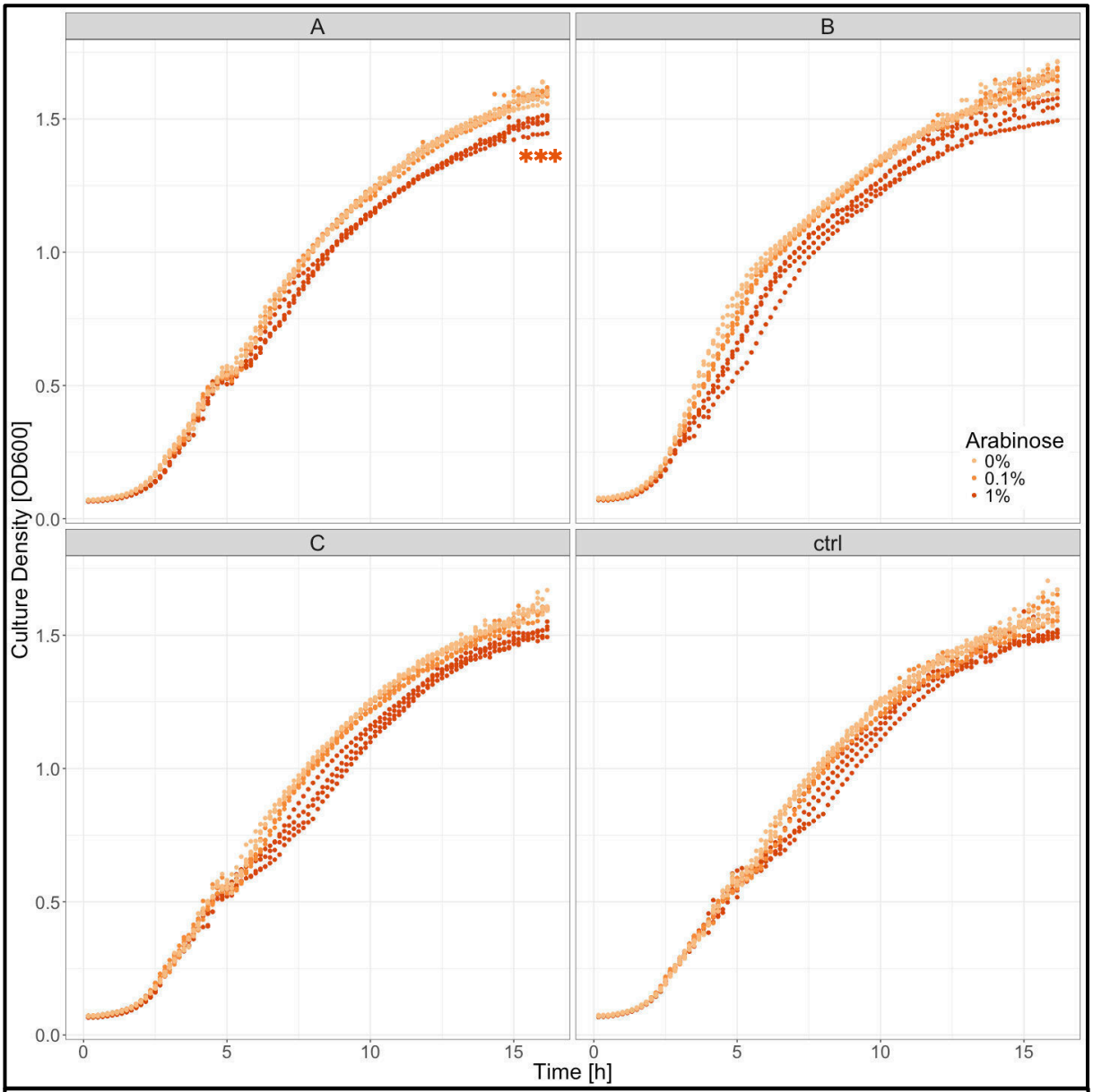


Figure III/6. Arabinose-adapted unmethylated Csy3::LacZ pHERD30T_SRM do not show reduced growth.

Culture density measurements as a proxy of bacterial growth at different arabinose concentrations over a course of 16 hours. All data points are shown. Strains shown are from adaptation experiment as detailed in Figure III/5 and were grown overnight without arabinose prior to this experiment. N = 4. ***p≈0.00027. P value refers to differences to 0% (w/v) arabinose treatment, see methods. All other differences are non-significant.

Therefore, to ensure methylation of Csy3::LacZ, we transformed cells with pHERD20S_SM, which confers streptomycin resistance and encodes HsdS and HsdM, but not HsdR. This incomplete RM system has the ability to methylate, but not to cleave DNA. pHERD20S_SM transformants were grown for 4 days in the

presence of 0.2% (w/v) arabinose to ensure complete methylation, followed by transformation with pHERD30T_SRM. Transforming methylated Csy3::LacZ with SRM in this way returned a higher EOT than transformation of unmethylated Csy3::LacZ at moderate arabinose concentrations (0.1-0.2% (w/v)) (Fig III/7).

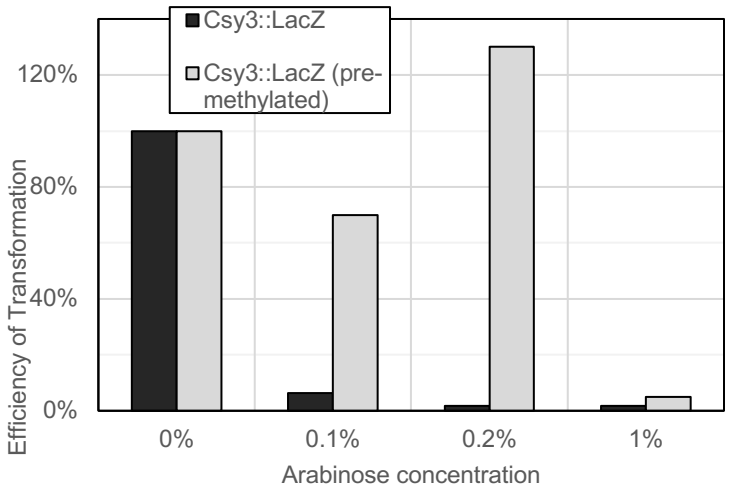


Figure III/7. Premethylation of Csy3::LacZ restores EOT of pHERD30T_SRM at moderate arabinose concentrations.

Efficiency of Transformation (EOT) of pHERD30T_SRM at different arabinose concentrations for unmethylated and premethylated (Csy3::LacZ pHERD20S_SM) bacteria. 100% EOT corresponds to colony forming units at 0% (w/v) arabinose, single replicate data shown.

To determine whether induction of the RM system bears the same costs as in unmethylated bacteria, we measured growth of premethylated pHERD30T_SRM transformants over 16 hours. Premethylated Csy3::LacZ pHERD30T_SRM does not show the same limited growth as seen in unmethylated transformants (Fig III/4 "previously methylated"). Together, these data show that premethylating the PA14 genome alleviates the toxicity of HsdR, presumably preventing self-DNA cleavage. In this way, the most drastic costs of carrying a novel RM system can be circumvented. This premethylation effectively simulates a lag before HsdR activity, which is the basis for various RA mechanisms (Prakash-Cheng and Ryu, 1993; Kulik and Bickle, 1996; Makovets, Doronina and Murray, 1999). Therefore, we might speculate such a RA mechanism is in place in *M. pulmonis* to ensure that the shufflon can switch specificity without causing toxicity.

In order to determine whether there are additional fitness costs associated with an established RM system, we directly compared growth of premethylated Csy3::LacZ pHERD30T_SRM with empty vector control at 0.1% (w/v) arabinose. Bacteria carrying the RM system grew slower, and reached a lower culture density after 16 hours (Fig III/8). This shows that even though initial RM toxicity can

be alleviated by premethylation, expressing RM proteins still bears a certain fitness cost.

Apart from these fitness costs, RM will also provide fitness benefits, such as phage resistance – for example, *Mycoplasma* phage P1 infectivity in a strain expressing only MpuUI is 500 times lower than when it infects a non-RM host (Dybvig, Sitaraman and French, 1998). To examine whether RM with the same specificity provides similar fitness benefits in the context of phage infection of *P. aeruginosa*, we exposed lawns of premethylated Csy3::LacZ pHERD30T_SRM transformants to four different *Pseudomonas* phages (DMS3vir, Φ1214, JBD5, LUZ24; Table S2). Intriguingly, we observed variation in our phenotypes between independent transformants for reasons that are currently unclear. In four of seven transformants, we observed partial immunity against some phages, mostly Φ1214 and JBD5 (Fig III/9). Three transformants reduced the EOP of these phages to ~0.5 compared to their empty vector control (as in transformant A in Fig III/9; other two transformants not shown), a single transformant (B in Fig III/9) reduced Φ1214 and JBD5 EOPs to ~0.2-0.5 compared to its empty vector

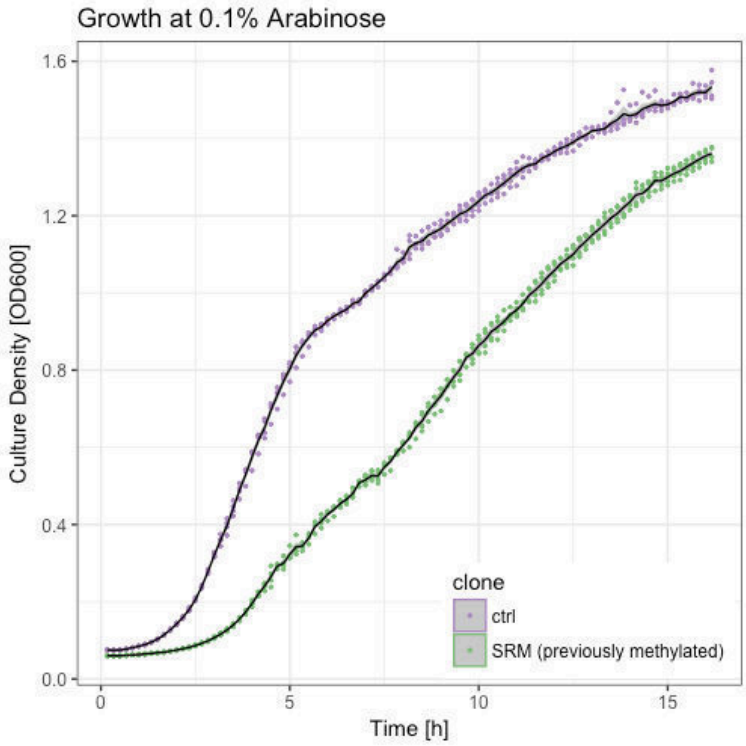


Figure III/8. Premethylated Csy3::LacZ pHERD_30T grows slower than bacteria without RM system.

Growth of Csy3::LacZ transformants over 16 hours at 0.1% (w/v) arabinose, as in Fig III/4. All data points are shown, with their average as a black line. Ctrl – Csy3::LacZ pHERD30T, SRM (previously methylated) – Csy3::LacZ pHERD30T_SRM (premethylated). N = 4. Intrinsic growth rate r estimated using Growthcurver (Sprouffske and Wagner 2016) and found to be significantly different by a T-test; $p=3.87x10^{-5}$.

control, and the final three transformants did not reduce phage EOPs (as in transformant C in Fig III/9; other two transformants not shown). This shows that RM provides partial protection against phages Φ1214 and JBD5, but is not sufficient as a stand-alone immune system.

This large discrepancy in phage restriction between MpuUI in *P. aeruginosa* (EOP=0.2-0.5) and MpuUI in *M. pulmonis* (EOP=0.002) could be due to several reasons. Firstly, *M. pulmonis* experiments were carried out with *Mycoplasma* phage P1. As we can see a difference in resistance levels to different *Pseudomonas* phages, phage P1 might simply have more RM recognition sites than any *Pseudomonas* phages trialled; restriction activity is known to be directly proportional to the number of restriction sites on the target (Wilson and Murray, 1991). The Mpu shufflon evolved in the background of *M. pulmonis*, therefore it might be adapted to restrict *Mycoplasma* phages with high efficiency. *Mycoplasma* phage P1 has a genome far smaller than all *Pseudomonas* phages used in this thesis (11.6 kb vs. 36-48 kb). Most Type I RM systems can be predicted to cleave

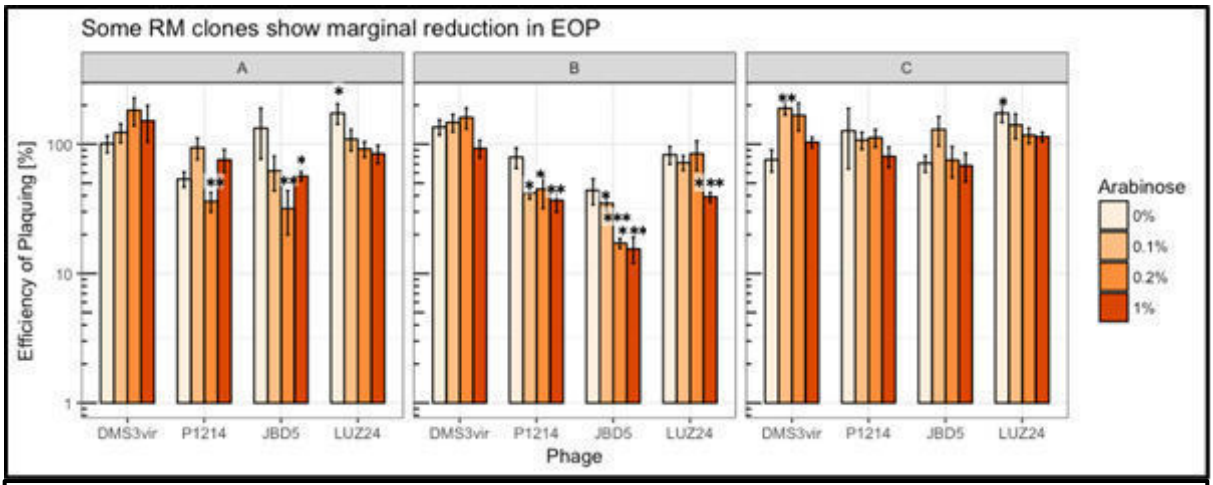


Figure III/9. pHERD30T_SRM confers low-level phage resistance to some premethylated Csy3::LacZ transformants.

Efficiency of Plaquing (EOP) as a measure of phage infectivity. Phages were spotted onto lawns of premethylated Csy3::LacZ transformants (A, B, C) at varying arabinose concentrations, and their number of plaques relative to their empty vector control at 0% (w/v) arabinose (not shown) recorded as EOP. Similar results as in A or C were obtained for two further strains each (not shown). Bars are means with standard errors; N = 4. Significance was tested by analysis of variances and post-hoc Tukey's multiple comparison of means. *p<0.05 (from left to right p≈0.013, p≈0.028, p≈0.014, p≈0.014, p≈0.035, p≈0.028); **p<0.01 (from left to right p≈0.0054, p≈0.005, p≈0.0053, p≈0.0044); ***p<0.001 (from left to right p≈0.0004, p≈3.3×10^{-5}, p≈0.00034)

a random DNA sequence of at least 33 kb in length (see above), phage P1's genome being considerably smaller may indicate that MpuUI targets conserved sequences that code for essential phage proteins. Alternatively, there might be as of yet undescribed *Mycoplasma* proteins that enhance activity of the RM system in its native host that are absent in PA14. In order to confirm whether this discrepancy is due to a smaller number of recognition sites in *Pseudomonas* phages, future experiments could determine the MpuUI receognition site by Csy3::LacZ pHERD30T_SRM methylome analysis (e.g. by PacBio sequencing (Rhoads and Au, 2015)) and map these onto phage genomes. Generally, RM immunity observed against *Pseudomonas* phages remains very limited and cannot be observed in all pHERD30T_SRM transformants. Nonetheless, the above experiments outlining *in vitro* activity, toxicity, and partial immunity together show that the *Mycoplasma* RM proteins can be active in PA14 and provide a basis for further investigations into RM effectivity, diversity, and interplay with CRISPR-Cas. Together, these results indicate that RM might not provide a great benefit, but show how RM carries significant costs, especially when first induced in an unmethylated host.

In summary, we found that introduction of the Mpu shufflon into CRISPR-KO *P. aeruginosa* is toxic. This toxicity can be circumvented by premethylation, where only the *hsdS* and *hsdM* subunits are introduced before transforming with the full RM system. These RM proteins are active within *P. aeruginosa*, and can cleave DNA *in vitro* as well as provide low-level phage resistance.
It remains unclear how fitness benefits can outweigh costs for shufflons if changing RM specificity bears the same costs as introducing an entirely novel RM system. *M. pulmonis* KD735-15, a strain without any active RM system, is referred to as "more stable" than *M. pulmonis* variants with RM activity (such as KD735-16), meaning KD735-15 is less likely to recombine *hsdS* pseudogenes to acquire RM activity (Dybvig, Sitaraman and French, 1998). This might be due to toxicity of RM induction even in its native *M. pulmonis*. Together, this could indicate that acquiring an entirely new RM system is very costly (KD735-15 is stable and does not recombine often to express RM; our data show high toxicity when inducing RM for the first time in *P. aeruginosa*), but introducing new specificity subunits is less costly and can occur more readily (KD735-16 strains are heterogeneous, indicating high frequency of specificity switches (Dybvig, Sitaraman and French,

1998)). One mechanism for this alleviation of toxicity when switching specificities could be the presence of more HsdS subunits: when switching specificity, “old” HsdS proteins will remain in the cell until they are degraded, free to bind HsdM and HsdR. This means that HsdR is more likely to be bound to a subunit which recognises a methylated sequence (the “old” HsdS protein) than in a scenario in which a bacterium acquires a new RM system, in which no HsdS proteins would be present.

In the future, we will be able to test this proposition by repeating the above experiments and trial toxicity with Csy3::LacZ pHERD30T_SRM transformed with a different SRM vector encoding alternative *hsdS* subunits, simulating a recombination event. If the costs of new specificity subunits remain high, it would mean that there is probably an additional mechanism for RA in place in shufflon-encoding bacteria.

Additionally, it will be interesting to see whether the benefits of RM will increase when trialling the RM system with alternative *hsdS* subunits – perhaps other specificities will be better suited to recognise *Pseudomonas* phages than MpuUI, and lead to a more drastic drop in EOP. If this is the case, RM benefits might only outweigh its costs for certain *hsdS* conformations, which would indicate that *M. pulmonis* and other bacteria encoding RM shufflons might favour certain specificity variants over others when exposed to phage.

Methods

in silico construct building

hsdS (MpuUI)*, hsdR,* and *hsdM* sequences of *hsd1* were found on Genbank (Clark *et al.*, 2016) under accession number L25415 (Dybvig & Yu 1994). The genes were separated and gene sequences were codon optimised for *Pseudomonas aeruginosa* PA14 using OPTIMIZER (Puigbò *et al.*, 2007). Where needed, unwanted restriction sites or predicted ribosome binding sites (RBS) within the optimised genes (identified using Prodigal (Hyatt *et al.*, 2010)) were removed by changing affected codons to the codon with the highest usage in PA01. After optimisation, protein sequences were verified by comparing the translation of codon optimisation (achieved with ExPASY (Artimo *et al.*, 2012)) with reported sequences using BLAST (Altschul *et al.*, 1990).

The T7 RBS together with surrounding nucleotides (32 nt stretch upstream of initiation codon) as found on pHERD-30T was introduced directly upstream of

each gene in non-coding regions. Further constructs were designed using DNA editing tools on Benchling (Benchling, 2015). For annotated sequences of all synthetic gene constructs see the supplemental data section at the end of this thesis.

Construct cloning

Construct synthesis and plasmid creation

The synthetic gene constructs were synthesised using ThermoFischer's GeneArt service. The synthesised SRM construct was subcloned into Novagen's pCDF-1b using NcoI and BamHI restriction sites. Afterwards, the construct was cloned into pHERD-30T using NcoI and HindIII restriction sites, generating pHERD30T_SRM. pHERD30T_SM was created by amplifying pHERD30T_SRM using primers S-KpnI-bwd and M-KpnI-fwd (Table S1). The product was re-ligated, resulting in a deletion of *hsdR*. To generate pHERD20S_SM, *hsdS* and *hsdM* were removed from pHERD30T_SRM and inserted into pHERD-20T using NcoI and HindIII sites. Subsequently, the streptomycin resistance gene from pCDF-1b was cut out and inserted upstream of pHERD-20T's multiple cloning site in a non-coding region using AgeI and SgrAI.

Restriction

NEB restriction enzymes NcoI-HF, KpnI-HF, BamHI-HF, HindIII-HF, SgrAI, and AgeI-HF were used as indicated. Approximately 1 µg of DNA was degraded in NEB CutSmart buffer using 10-20 U of each enzyme and topped up to a total reaction volume of 50 µl with nuclease-free distilled water. Restriction reactions were incubated at 37°C for 1-2 hours and successful restriction was verified by comparing to a no-restriction-enzyme control reaction under the same conditions using gel electrophoresis.

Vector de-phosphorylation and gel extraction

Without a clean-up of restriction reactions, restricted vectors were de-phosphorylised in 60 µl volumes using 5 U of NEB Antarctic Phosphatase and its supplied buffer. De-phosphorylation reactions were carried out at 37°C for one hour. Desired restriction fragments were isolated using Qiagen's Gel Extraction Kit per the manufacturer's instructions.

Plasmid ligation

Desired fragments were inserted into the appropriate vectors in 20 µl reactions using 200 U of NEB T4 ligase and its supplied buffer. Amounts of vector and insert for ligation reactions were calculated using a ligation calculator tool by the

University of Düsseldorf (Insilico, 2017). Reactions were carried out for 10 minutes at room temperature and, if no successful ligation occurred, at 4°C overnight.

Expression in *E. coli*, plasmid extraction, and cloning verification

Ligated plasmids were expressed in NEB 5-α electrocompetent *E. coli* per the manufacturer's instructions. Amplified plasmids were extracted from liquid cultures derived from a single colony using ThermoFischer's GeneJET Plasmid Miniprep Kit as per the manufacturer's instructions. A successful cloning process was verified by restriction fragment analysis (restriction reactions as above followed by visualisation using gel electrophoresis) and by PCR or Sanger sequencing where appropriate.

pHERD30T_SRM amplification

SM was amplified using a high-fidelity recombinant Pfu polymerase from Thermo Scientific. The amplification reaction was carried out in a total volume of 50 µl with 2 ng template plasmid DNA, 1 µM of each primer, 0.2 µM of each dNTP, and 2.5 U Pfu polymerase in the supplied buffer. The reaction was cycled as recommended in Thermo Scientific's documentation. Forward and reverse primers amplified SRM-30T in such a way that the entire plasmid bar most of the *hsdR* gene was copied, both primers contain KpnI restriction sites with an additional three-nucleotide overhang (Table S1). Therefore, the PCR-product was digested and then ligated to be re-circularised as described above.

In vitro analysis

Pseudomonas transformation

Pseudomonas aeruginosa PA14 strains were made electrocompetent by pelleting 1-1.5 ml of an overnight culture in LB. The pellet was washed twice with 1 ml of a 300 mM sucrose solution, and resuspended in 100 µl sucrose.

300-500 ng of plasmid DNA were added to the resuspended pellet and mixed by flicking. These were electroporated at 2.5 kV for 3-5 ms in 1 mm gap cuvettes. Immediately, 1 ml room temperature LB was added and the culture resuspended in this, taking care not to damage the cells. After a shaking incubation at 37°C for 0.5 - 2 hours, 50 - 100 µl of undiluted cultures were plated onto LB with appropriate antibiotics (50 µg/ml gentamycin or streptomycin) and incubated overnight at

37°C. Controls included cells transformed with H_2O rather than DNA, resulting in no colonies on antibiotics plates.

Protein lysate extraction

Csy3::LacZ pHERD30T_SRM was grown overnight at 37°C in 15 ml LB + 50 µg/ml gentamycin (GM50). Arabinose was added to a total concentration of 1% (w/v), and the bacteria were incubated at 28°C for 3 hours before being centrifuged (15 min, 3500 rpm). Culture pellets were resuspended in 10 ml Tris buffer (20 mM Tris, 150 mM NaCl, 1mM DTT). Resuspended pellets were sonicated (10 second pulse/rest cycle) for 1-2 minutes and centrifuged. The supernatant was filtered through a 0.45 µm filter and the lysate stored at -80°C in 40% (w/v) Glycerol. Lysate was extracted in two replicates.

To restrict, 1 µg PA14 WT gDNA (extracted using QIAamp DNA Kit as per the manufacturer's instructions) or 300 ng Φ1214 gDNA (sample kindly provided by Sean Meaden; extracted using Norgen Biotek phage DNA isolation kit as per the manufacturer's intructions) was digested in 0.44 x TMD buffer (50 mM Tris, 10 mM MgCl2, 1 mM DTT, pH = 8.0) together with 4mM ATP (pH = 7.0) using 5 µl of lysate (or H2O as control) in a total volume of 20 µl. The reaction was incubated at 37° overnight (~15 hours).

DNA was separated using electrophoresis on a 1% (w/v) agarose gel and visualised using RedSafe.

Construct toxicity

Growth curves

Overnight cultures of bacteria as indicated were diluted 1:100 into LB + GM50 containing 0%, 0.1%, and 1% (w/v) Arabinose. Diluted cultures were aliquoted into a 96-well plate with 4 replicates each. Overnight, a measurement was taken by a Biotek Synergy 2 plate reader, on a protocol that shook the plate at the setting "slow" at 37°C while taking an OD600 measurement of every well every 10 minutes for 16 hours.

Adaptation Experiment

Csy3::LacZ pHERD30T_SRM cultures (not pre-exposed to arabinose) were grown in LB + GM50 overnight in three replicates (T0), and then transferred into LB + GM50 + 0.001% (w/v) arabinose (T1). T1 bacteria were grown overnight and then transferred into LB + GM50 + 0.01% (w/v) Arabinose (T2). In the same way, further timepoints included T3 at 0.1% (w/v) Arabinose and T4 at 1% (w/v)

Arabinose. As a control, Csy3::LacZ pHERD30T was transferred in the same way. Cultures of every timepoint were plated out onto LB + GM50 + 1% (w/v) arabinose in a 10^{-5} dilution. Colony forming units (CFU) were counted and the growth calculated by (CFU for Csy3::LacZ pHERD30T_SRM) / (CFU for Csy3::LacZ pHERD30T) at the respective timepoint (Fig III/5a).

Csy3::LacZ was transformed with pHERD30T, ancestral pHERD30T_SRM, or pHERD30T_SRM isolated from T4 cells in 4-5 replicates. The same transformation mixture was plated on LB + GM50 plates as well as on LB + GM50 + 1% (w/v) arabinose plates in equal volumes. Controls included H_2O - transformed bacteria as well as transformants plated on LB. Colonies on all plates were counted. Efficiency of Transformation (EOT) was calculated as (number of colonies on 1% (w/v) arabinose) / (number of colonies on 0% (w/v) arabinose).

Phage infectivity

Phage infectivity was measured by carrying out spot assays: Appropriate strains were grown overnight as indicated, then 300 - 600 µl of culture were mixed with 12.5 ml LB broth containing 0.5% (w/v) Agar and antibiotics/arabinose as indicated and plated onto a square plate with LB agar. Phages were diluted from 10^0 - 10^{-7} in M9 salts (0.6% $Na_2HPO_4 \cdot 7H_2O$, 0.3% KH_2PO_4, 0.05% NaCl, 0.1% NH_4Cl; all w/v) and 5 µl of each dilution spotted onto bacterial lawns.

After overnight incubation at 37°C, apparent titres could be calculated by counting plaques at appropriate dilutions. Efficiency of Plaquing (EOP) was calculated as a measure of phage infectivity by (titre on Csy3::LacZ pHERD30T_SRM strain) / (titre on empty vector control) as indicated.

Statistical analyses

Statistical analyses were carried out using R software (R Core Team, 2017), specific tests used for each experiment are highlighted in figure legends. For bacterial growth curves (Figs III/4 and III/6), the package Growthcurver (Sprouffske and Wagner, 2016) was used to estimate various growth curve statistics. As a proxy for bacterial growth, carrying capacity k of bacterial growth at moderate (0.1% w/v) and high (1% w/v) arabinose concentrations for each strain was compared to k for its respective growth in the absence of arabinose and their difference assessed with an analysis of variances and a Tukey's post-hoc test.

Restriction-Modification and CRISPR act synergistically to provide high levels of phage resistance

Abstract

The interplay of the two most common bacterial immune systems, Restriction-Modification (RM) and CRISPR-Cas remains insufficiently investigated. We transformed *Pseudomonas aeruginosa* PA14 with an artificially encoded RM system to generate a strain with both a Type I-F CRISPR-Cas and a Type I RM system. Together, these immune systems provide complete resistance against all pilus-specific bacteriophages tested, and reduce infections of LPS-specific LMA2 by 2-3 orders of magnitude. Some transformants only show partial resistance, perhaps due to incomplete RM expression. When transforming different CRISPR-knockout PA14 strains, resistance cannot be detected, indicating CRISPR-Cas is an essential component for this resistance. CRISPR spacer acquisition can be detected after coevolution with DMS3vir, but it is unlikely that RM aiding spacer acquisition is the mechanism behind this resistance. More likely, RM methylation patterns regulate genes responsible for resistance in a CRISPR-dependent manner, as the restriction endonuclease HsdR is not essential for the effect of joint resistance. As this model leaves some data unexplained, the mechanism of joint RM and CRISPR resistance remains to be investigated.

Introduction

Bacterial growth in many environments is limited by their natural predator, the bacteriophage (phage). Bacteria have evolved sophisticated defense mechanisms against phage and other predatory genetic elements. Restriction-Modification is a nearly ubiquitous (>90% of bacteria encode RM-systems (Stern and Sorek, 2011)) innate bacterial immune response that functions by self-recognition through methylation patterns and cleavage of non-self DNA, as discussed in detail in the previous chapters. CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats - CRISPR associated) on the other hand is a more recently discovered adaptive bacterial immune response, but is nonetheless common throughout bacteria as well as archaea (50 and 80% respectively (Grissa, Vergnaud and Pourcel, 2007b)).

CRISPR-Cas adaptation functions by integrating short segments of phage genome (protospacers) into the CRISPR locus as spacers (Modell, Jiang and Marraffini, 2017). In a process known as expression, these CRISPR loci are transcribed as pre-CRISPR RNA (pre-crRNA) and processed into short crRNAs that are composed of a single spacer sequence flanked by partial repeats (Brouns *et al.*, 2008). In type I CRISPR-Cas systems, these crRNA molecules act as a guide for target DNA recognition by the Cascade complex, which is a ribonucleoprotein complex comprised of several Cas proteins and crRNA (Makarova *et al.*, 2015). Upon re-infection by the same phage, the Cascade complex specifically binds phage DNA through base-pairing between the crRNA and the protospacer sequence (Wiedenheft *et al.*, 2011). Subsequent recruitment of the nuclease Cas3 causes cleavage of the target DNA in a process known as CRISPR interference (Westra *et al.*, 2012). Our model organism *Pseudomonas aeruginosa* PA14 possesses a Type I-F CRISPR-Cas system with two CRISPR loci, the second of which is highly active (Westra *et al.*, 2015). Its Cas1 nuclease is active in the adaptation step, whilst Cas2 and Cas3 are fused (Makarova *et al.*, 2015). For interference, Cas2-3 are recruited to the Cascade complex, which is formed by Csy1-4 and crRNA (Wiedenheft *et al.*, 2011).

To date, there have been limited investigations into the direct interplay between CRISPR-Cas and RM, even though they are often coexpressed in bacterial hosts. Immunity genes such as these tend to cluster in defense islands within the bacterial genome, which may be an indication of functional coupling (Makarova, Wolf and Koonin, 2013). In a *Streptococcus thermophilus* model, RM and CRISPR were shown to increase phage resistance when combined (Dupuis *et al.*, 2013). Building on this study, it was found that the presence of defective phage particles aid in CRISPR spacer acquisition. RM-deactivated phage particles were most effective in promoting spacer acquisition, indicating there might be synergistic interactions (where the combinatory effect is greater than the sum of parts) between the two immune systems (Hynes, Villion and Moineau, 2014). In *Enterococcus faecalis*, CRISPR and RM form a non-synergistic barrier to horizontal gene transfer by plasmids (Price *et al.*, 2016). All three studies described interactions of a Type II RM and a Type II CRISPR-Cas system. Another level of complexity is added to this through gene regulation by methylation – RM systems can up- or downregulate genes due to their methylation activity, making it more

difficult to tease apart mechanistic basis of synergistic effects (i.e. direct interactions as opposed to gene regulatory epigenetic effects). Furthermore, in some cases CRISPR may play a role in gene regulation (Westra, Buckling and Fineran, 2014).

For the model organism *P. aeruginosa* PA14, bacteria-phage dynamics in the presence of CRISPR-Cas are well-investigated (van Houte *et al.*, 2016). In the previous chapter, we established a model system to study CRISPR-RM dynamics by adapting a Type I RM shufflon found in *Mycoplasma pulmonis* for expression in PA14, and we further investigated bacteria-phage dynamics in the presence of RM but absence of CRISPR-Cas. Therefore, in this chapter we aim to investigate the interplay of the Mpu RM-system and a Type I-F CRISPR-Cas in a *P. aeruginosa* host, and determine the combined effect of these immune systems on phage resistance.

Results and Discussion

Our RM system of choice is derived from a Type I RM shufflon found in *M. pulmonis*, and was adapted for expression in PA14 by cloning the RM genes into an expression plasmid under control of an arabinose-inducible promoter (Table S1). Initially, we transformed PA14 WT with pHERD20S_SM (which encodes RM without HsdR, ensuring restriction cannot occur) and induced construct expression with arabinose to achieve genome methylation. These strains were then transformed with pHERD30T_SRM to introduce the entire RM system. We isolated 6 premethylated transformants from two independent transformations (hereafter called WT pHERD30T_SRM A-F, or transformants A-F for brevity).

Initially, we tested infectivity of pilus-specific *Pseudomonas* phages DMS3vir and JBD5, and of LPS-specific LMA2 against bacteria co-expressing CRISPR and RM (note that bacteria do not have *a priori* CRISPR resistance, i.e. bacteria lack CRISPR spacers targeting phages, but the presence of the CRISPR-Cas genes allows bacteria to acquire CRISPR-based phage resistance in response to infection). To these ends, we measured Efficiency of Plaquing (EOP) of these phages on lawns of WT pHERD30T_SRM transformants compared to empty vector controls (which therefore express CRISPR but not RM) at 0.2% (w/v) arabinose in 5-6 replicates. We found that all transformants show resistance against all phages with two distinct phenotypes of resistance (Fig IV/1a): Transformants B and E

were found to have a “partial-resistance” phenotype, in which LMA2 EOP were reduced by 1-2 orders of magnitude (EOP ~0.03-0.05) compared to the empty vector control. JBD5 and DMS3vir titres were difficult to count because plaques were faint, estimated EOPs for these phages were 2-3 orders of magnitude lower (EOP ~0.002-0.004) on transformants B and E compared to the control. All other WT pHERD30T_SRM transformants showed a “complete-resistance” phenotype, in which titres of DMS3vir and JBD5 on pHERD30T_SRM-transformed bacteria remained below the limit of detection (~200 pfu/ml). LMA2 retained the ability to infect these strains, albeit with an EOP reduced by 2-3 orders of magnitude compared to the empty vector control (EOP ~0.005-0.007). Despite its infection only

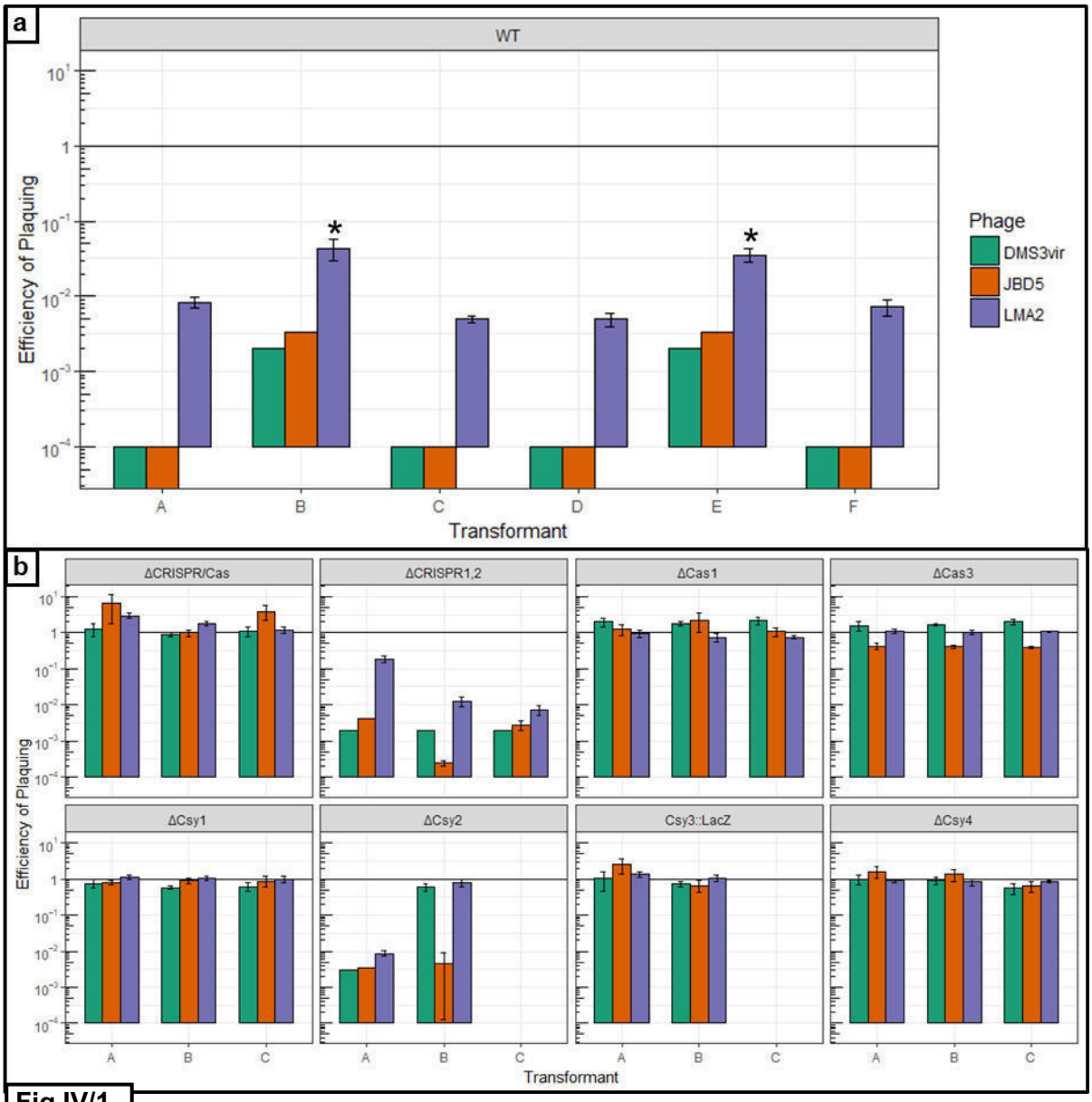


Fig IV/1

Figure IV/1. Coexpression of CRISPR and RM leads to phage resistance.
Efficiency of plaquing (EOP) of DMS3vir, JBD5, and LMA2 on lawns of PA14 WT (a) and CRISPR-knockout (b) pHERD30T_SRM transformants compared to their respective empty vector controls (black line at EOP=1). Bars represent means and their standard error, N=5-6. Bars extending downwards signify EOP below the limit of detection. **a** Most transformants (A, C, D, F) have a "complete-resistance" phenotype and are completely resistant against DMS3vir and JBD5. Transformants B and E have a "partial-resistance" phenotype and are partially resistant against all phages tested. * $p<0.05$. Significance tested with analysis of variances and post-hoc Tukey's HSD. EOPs of LMA2 on B and E are significantly higher than EOPs on A, C, D, F (difference between A and E is nonsignificant at $p\approx0.068$, between E and F is nonsignificant at $p\approx0.052$. B-A $p\approx0.0095$, B-C $p\approx0.0040$, B-D $p\approx0.0040$, B-F $p\approx0.0070$, E-C $p\approx0.031$, E-D $p\approx0.031$). **b** No CRISPR-knockout transformant has a complete-resistance phenotype. EOP on most transformants does not change compared to their empty vector control, ΔCRISPR1,2 and ΔCsy2 show levels of resistance similar to partial-resistance phenotype transformants in (a). Missing values for transformants C of ΔCsy2 and Csy3::LacZ are due to bacterial lawns too faint to enumerate plaques, their phenotype is scored in Table IV/1.

being partially blocked, LMA2 EOPs remained significantly higher on partial-resistance WT pHERD30T_SRM transformants compared to complete-resistance transformants. These data reveal that co-expressed CRISPR-Cas and RM provide *P. aeruginosa* with phage resistance to a higher degree than from either CRISPR-Cas (empty vector control) or RM (Chapter III) alone. Additionally, this result uncovers a certain degree of variation in phage resistance (Fig IV/1a, Table IV/1 WT); perhaps the two transformants with partial resistance express their RM system at a lower level than transformants with complete resistance. Future reverse-transcription qPCR of *hsd* RNA in these transformants would help to confirm whether this is the case. Furthermore, JBD5 encodes anti-CRISPR (Acr) proteins that block Type I-E as well as Type I-F systems (Pawluk *et al.*, 2014), therefore WT pHERD30T_SRM resistance against this phage indicates that resistance is not mediated through a classical CRISPR immune response.

To further explore this question, we first examined whether all components of the CRISPR-Cas immune response are essential to this synergistic resistance phenotype of CRISPR-Cas and RM. We transformed several PA14 CRISPR knockout strains (ΔCRISPR1,2, ΔCRISPR/Cas, ΔCas1, ΔCas3, ΔCsy1, ΔCsy2, Csy3::LacZ, and ΔCsy4) with pHERD30T_SRM (by first premethylating strains with pHERD20S_SM as described previously) and isolated three transformants from each transformation. EOP of DMS3vir, JBD5, and LMA2 compared to their infectivity on a corresponding empty-vector control was determined at 0.2% (w/v) arabinose in 5-6 replicates for each transformant. Strikingly, nearly all CRISPR-mutants that were transformed with the RM system were susceptible to phage and show levels of infection similar to their empty vector controls (Fig IV/1b).

Strain	**Transformant**					
WT	A	B	C	D	E	F
ΔCR/Cas	A	B	C			
ΔCRISPR1,2	A	B	C			
ΔCas1	A	B	C			
ΔCas3	A	B	C			
ΔCsy1	A	B	C			
ΔCsy2	A	B	C			
Csy3::LacZ	A	B	C			
ΔCsy4	A	B	C			

Table IV/1. Phenotypes of pHERD30T_SRM transformants.
pHERD30T_SRM transformants of several PA14 strains were tested for resistance to DMS3vir, JBD5, and LMA2. Green-susceptible (only low-level or no resistance to any phage), Orange-"partial-resistance" phenotype (partially resistant, decreasing EOPs to ~0.001-0.002 for DMS3vir and JBD5 and to ~0.008-0.05 for LMA2), Red-Complete-resistance phenotype (entirely resistant to DMS3vir and JBD5, decreasing LMA2 EOP to ~0.01). Phenotypes were determined by spot assays, N = 5-6 per phage. See text for more detailed description of phenotypes.

All three ΔCRISPR1/2 pHERD30T_SRM transformants (lacking both CRISPR loci) as well as ΔCsy2 pHERD30T_SRM A showed more phage resistance than their empty vector control, in a pattern reminiscent of the partial-resistance phenotype observed in WT transformants. In summary, these data clearly indicate that presence of all CRISPR-Cas components is necessary for the complete-resistance phenotype.

As before, some transformants showed a partial-resistance phenotype, including transformants of PA14 mutants that lack both CRISPR loci or the Csy2 gene (Table IV/1, Figure IV/1b). We believe that this variation in the levels of resistance is likely due to variation in RM expression levels between different transformants. Perhaps partial resistance is observed when the MpuUI RM system is fully functional, and full sensitivity is observed when mutations are acquired in the RM genes, which we demonstrated can be rapidly selected for because of the large fitness costs of RM (see Chapter 3). This could indicate that constraints of successfully introducing a functional RM system are larger than assumed in the previous chapter. Therefore, future experiments should examine in more detail whether CRISPR-knockout pHERD30T_SRM transformants without any apparent resistance encode functional RM systems. This could for example be done by sequencing pHERD30T_SRM isolated from CRISPR-knockout transformants. Alternatively, we could test the RM activity of isolated pHERD30T_SRM *in vitro* using the assays described in Chapter 3 (see Fig. III/2) as well as *in vivo* (by transforming strains with plasmids containing MpuUI's recognition sites, which are yet to be identified (see Chapter 3)). Transformants of each CRISPR mutant were isolated from the same transformed population, therefore transformants are likely to be genetically identical, hence explaining why replicate transformants of the same mutant all have the same resistance phenotype (with few exceptions). Collectively, the data above are consistent with the idea that the full resistance phenotype is due to a CRISPR-mediated immune response that is facilitated by the presence of the RM system. However, as mentioned above, this appears inconsistent with the fact that one of the phages encodes an Acr, which would be expected to reduce or eliminate such a CRISPR-dependent synergistic effect on phage resistance.

To probe this in more detail, we directly measured the evolution of CRISPR-mediated resistance in bacteria encoding both CRIPSR and RM, and in bacteria encoding CRISPR only. A previous study suggested that synergistic resistance of RM and CRISPR-Cas may be due to RM enhancing spacer acquisition (Hynes, Villion and Moineau, 2014), and observing a lack of phage resistance in ΔCas1 pHERD30T_SRM transformants (Cas1 is the nuclease responsible for spacer acquisition; Fig IV/1b) may corroborate this hypothesis. Therefore, we allowed WT pHERD30T_SRM transformants to evolve together with DMS3vir or JBD5 to

transformants as well as a no-phage control for all treatments. Every day, the growing cultures were transferred into fresh LB + 50 µg/ml gentamycin (GM50) + 0.2% (w/v) arabinose medium, and 10^9 pfu phage were added, giving us an initial MOI (multiplicity of infection) of ~20. As expected, we could not detect any spacer acquisition after three days in any of the control treatments (in LB and with a high MOI, CRISPR-mediated resistance does not evolve (Westra *et al.*, 2015)). In the

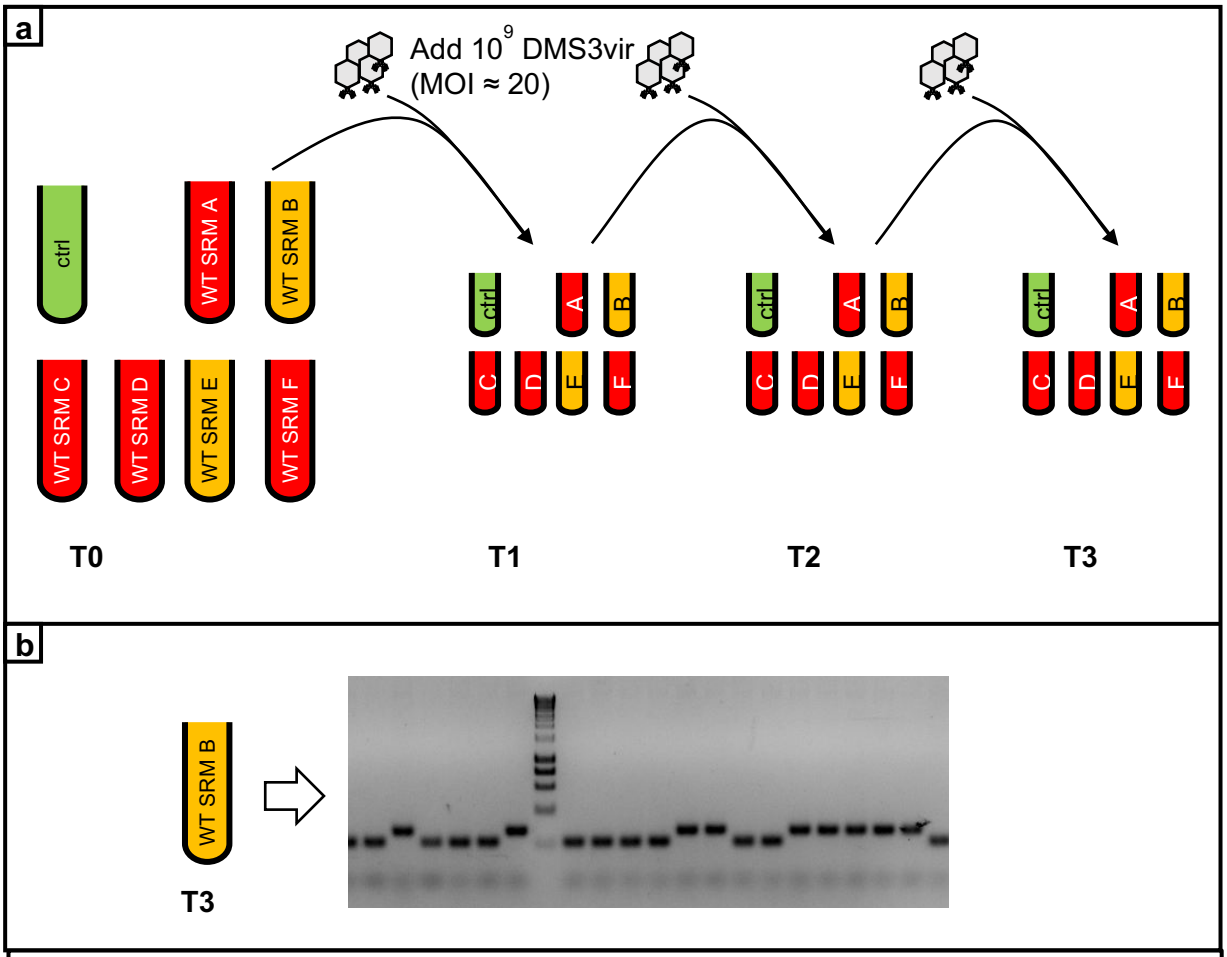


Figure IV/2. RM expression can promote CRISPR spacer acquisition.
a Representation of 3-day bacterial evolution together with DMS3vir. Strains were transferred into fresh medium daily and 10^9 DMS3vir or JBD5 (not shown) added with each transfer. Additional controls (not shown) included Csy3::LacZ pHERD30T_SRM and empty vector transformants, as well as a no-phage control for each treatment. Colours represent phenotypes. Red: complete-resistance, orange: partial-resistance, green: susceptible (Table IV/1). WT SRM – WT pHERD30T_SRM transformant; ctrl – WT pHERD30T empty vector control. **b** After 3 days, spacer acquisition was surveyed in all treatments and replicates by PCR of CRISPR loci. Only WT SRM B showed spacer acquisition (10/21 clones screened had acquired a spacer in CRISPR2), shown is a representative image of a 2% (w/v) agarose gel.

only instance where spacer acquisition was observed, we found that nearly half (10/21) of clones of WT pHERD30T_SRM B treated with DMS3vir had acquired a spacer in the CRISPR2 locus (Fig IV/2). These data show that the co-expression of CRISPR-Cas and RM can promote CRISPR spacer acquisition. However, as this was only observed in one transformant with a partial-resistance phenotype, enhanced spacer acquisition cannot explain the complete-resistance phenotype seen in most transformants.

Therefore, we hypothesised that complete-resistance phenotypes observed may be due to gene regulatory effects of RM and CRISPR coexpression. In particular, the combined activity of these immune systems may cause downregulation of the pilus, because we observed higher resistance of WT pHERD30T_SRM transformants against pilus-specific phages than against LMA2.

To examine whether the mechanism of RM-CRISPR synergistic resistance may be pilus downregulation, we investigated bacterial streaks of WT pHERD30T_SRM A for a visible change in phenotype, as can readily be observed for pilus mutants of PA14 which causes a small colony morphology (Westra *et al.*, 2015). We found that streaks for both WT pHERD30T_SRM as well as WT pHERD30T_SM resembled pilus mutants of PA14, whilst the morphology of empty vector and CRISPR-knockout (Csy3::LacZ) controls resembled that of untransformed PA14 Csy3::LacZ (Fig IV/3a). These data confirm that co-expression of RM and CRISPR could cause loss of pilus, resulting in resistance to pilus-specific phages. Intriguingly, RM causes a change in morphology in the WT PA14 strain even in the absence of the HsdR gene (WT pHERD30T_SM). This indicates that resistance may be dependent on methylation rather than restriction activity, which is consistent with RM regulating gene expression. Furthermore, these data suggest that the resistance phenotype is constitutively expressed, since the altered colony morphology was also observed in absence of phage.

To further explore putative pilus loss, we assayed phage adsorption which would be predicted to be reduced when bacteria have a lower pilus expression. To these means, we infected WT pHERD30T_SRM, WT pHERD30T_SM, and an empty-vector control with Φ1214 and determined the titre of phage in free solution after eight minutes of incubation for an estimate of percentage of phage bound to bacteria. As an additional control, we incubated phage in the absence of bacteria.

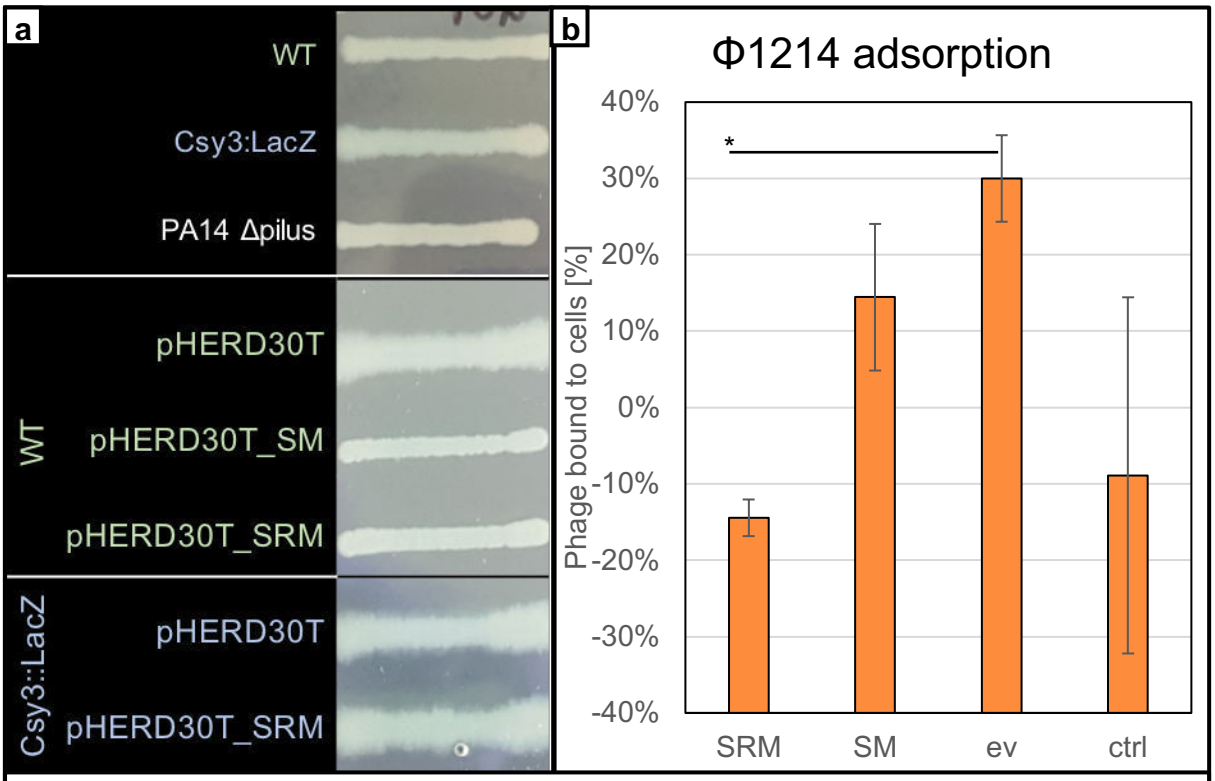


Figure IV/3. WT pHERD30T_SRM transformants resemble surface mutants.
a Streaks of different transformants. Untransformed WT, Csy3::LacZ, and PA14 Δpilus (pilus mutant) were streaked onto LB; all other strains onto LB + GM50 + 0.2% (w/v) arabinose and incubated for ~24 hours **b** Phage adsorption assay. Φ1214 was extracted after incubation with different WT transformants for 8 minutes. Percentage of phage bound to cells was calculated by comparing extracted phage titres with titres of phage initially added. Bars show means with their standard error; n = 3. SRM: WT pHERD30T_SRM, SM: WT pHERD30T_SM, ev: WT pHERD30T, ctrl: no-bacteria control. Significance tested by T-test of means between SRM and ev, *p<0.05 (p≈0.013)

We found that WT pHERD30T_SRM A showed a significant reduction in phage adsorption compared to the empty vector control (p ≈ 0.013; Fig IV/3b). Therefore, these results further corroborate the hypothesis that RM and CRISPR-Cas coexpression induces phage resistance through loss of the pilus. As the synergistic resistance effect of CRISPR and RM may be possible without HsdR (Fig IV/3a WT pHERD30T_SM), perhaps resistance is due to CRISPR-dependent RM gene regulation through methylation.

To confirm whether the complete-resistance phenotype can be observed in absence of HsdR, we tested infectivity of an assortment of phages (DMS3vir, Φ1214, JBD5, LUZ24, Φ68, JBD18, and JBD25) on lawns of WT pHERD20S_SM and WT pHERD30T_SRM A. As controls, we infected untransformed and empty vector control bacteria as well as the corresponding CRISPR-knockout Csy3::LacZ pHERD30T_SRM and pHERD20S_SM transformants. We found that, while control bacteria showed similar levels of infection with relatively small variations (not shown), EOP on both WT pHERD30T_SRM and WT pHERD20S_SM was below the limit of detection (~ 200 pfu/ml; Fig IV/4) for all *Pseudomonas* phages tested. This resistance was very consistently observed in more than five repeat experiments, and no phage infections could be observed when testing infectivity with plaque assays (not shown) which have a lower limit of detection of ~100 pfu/ml. These data confirm that HsdR is not an essential component for the complete-resistance phenotype, which may therefore be caused by methylation patterns. Additionally, the repertoire of phages against which *P. aeruginosa* can become entirely resistant is expanded by Φ1214, LUZ24, Φ68, JBD18, and JBD25, all of which are known to bind to *P. aeruginosa*'s pilus (LUZ24's receptor is unknown, but experiments in our lab show that it cannot infect a PA14 surface mutant lacking the pilus).

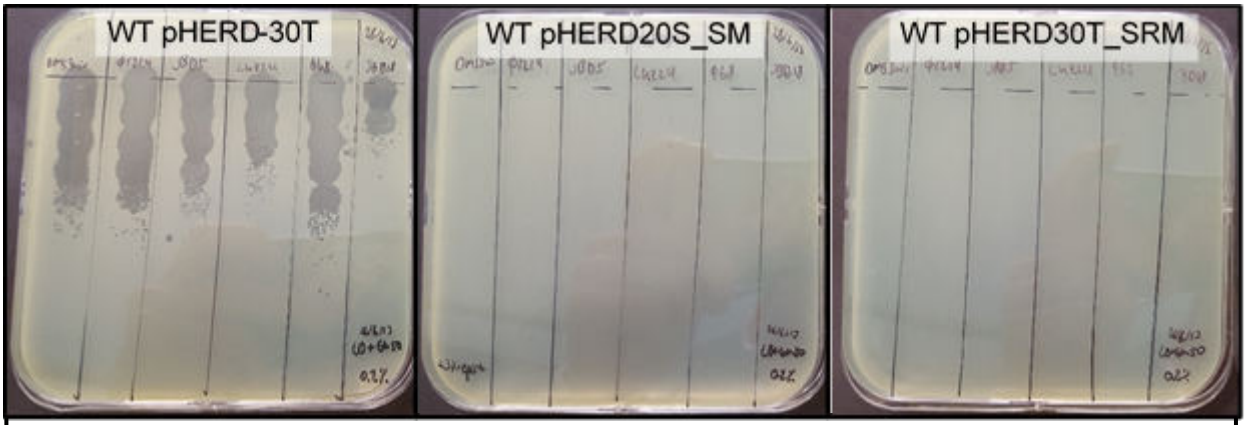


Figure IV/4. HsdR is not essential for synergistic RM and CRISPR resistance. Spot assays of phage infecting lawns of PA14 WT transformants. Phage infection remains entirely under the level of detection for WT RM-transformants. Pictures of representative spot assays at 0.2% (w/v) arabinose comparing empty vector control to pHERD30T_SRM/pHERD30T_SM transformants. Phages from left-right: DMS3vir, Φ1214, JBD5, LUZ24, Φ68, JBD18. Identical results were obtained in at least 4 repeat experiments and with phage JBD25.

In our model for CRISPR- and RM-dependent downregulation of pilus expression, RM methylates the CRISPR locus, enhancing its expression and constitutively activating it. In turn, activated Cas proteins will use crRNA derived from pre-existing spacers as guides to its own genome. With a partial match between crRNA and genome, CRISPR-Cas can have gene-regulatory rather than nucleolytic activities (Zegans *et al.*, 2009; Westra, Buckling and Fineran, 2014). Therefore, to test whether there is capacity for pilus gene regulation through CRISPR-Cas to occur, we analysed the PA14 genome for partial matches to spacers present in either PA14 CRISPR locus. For more relevant matches, we restricted the search to sequences that exactly match the spacers' seed sequence and found 27 partial matches across the PA14 genome. Three matches additionally fulfilled PAM (proto-spacer adjacent motif) requirements to be targeted by CRISPR (Table IV/2a). These include a hypothetical protein with unknown function, an autotransporter domain-containing esterase, and *serB*, a gene coding for a phosphoserine phosphatase that plays a role in serine biosynthesis. When removing the restriction for PAM complementarity, other metabolic genes and some DNA binding proteins can be targeted, too (Table IV/2b). In conclusion, while CRISPR cannot directly target a sequence in or near a pilus gene to regulate its expression, there are other viable targets present that might affect pilus expression further downstream. To identify whether this is the case, we would have to investigate whether a knockout or overexpression of *serB* or other targets can modulate pilus expression. Furthermore, we would have to find the recognition sequence of MpuUI to determine whether RM can enhance CRISPR expression in the first place.

During the above bioinformatics analyses, we also identified a partial match within *hsdR* on pHERD30T_SRM to CRISPR spacer 2.18 (match fulfils seed restriction, but not PAM; not shown) that may enable CRISPR to modulate *hsdR* expression, which could result in lower toxicity of RM induction in a host expressing CRISPR than observed in CRISPR-deficient hosts in the previous chapter. Therefore, we investigated toxicity of pHERD30T_SRM to unmethylated WT transformants to determine whether induction of a novel RM system is associated with the same toxicity as in a CRISPR-knockout background. To these means, we tracked growth of unmethylated WT pHERD30T_SRM and WT pHERD30T_SM while inducing RM expression with arabinose. While RM induction limited WT pHERD30T_SRM growth compared to its empty vector control to

a Partial matches in PA14 genome fulfilling seed and PAM restrictions			
Spacer	**Match**	**Gene**	**Product**
2.21	6026892-6026878	PA14_RS27540	autotransporter domain-containing esterase
1.2	3729879-3729853	PA14_RS16950	Hypothetical protein
1.2	5840199-5840175	*serB*	Phosphoserine phosphatase
b Partial matches in PA14 genome fulfilling only seed restriction			
2.2	5881986-5881959	PA14_RS26955	acyl-CoA dehydrogenase
2.18	1461715-1461689	PA14_RS06835	Bifunctional uridylyltransferase/uridylyl-removing protein
1.2	3207203-3207176	PA14_RS14650	MCE family protein
1.2	5533142-5533115	PA14_RS25340	MCE family protein
1.2	5451699-5451670	PA14_RS24975	hypothetical protein
1.2	3365462-3365486	PA14_RS15350	carbamoyltransferase
2.6	4454017-4454038	PA14_RS20355	flagellar assembly protein FliH. Partial match sequence is 98nt upstream of *fliI* (a flagellum-specific ATP synthase).
2.6	567437-567409	PA14_RS02600	allophanate hydrolase
2.17	1091372-1091396	PA14_RS05130	cysteine hydrolase
2.17	5174166-5174141	PA14_RS23705	septum formation inhibitor Maf
2.17	4784617-4784599	*acnD*	Fe/S dependent 2-methylisocitrate dehydratase
2.4	108344-108372	*clpV*	type VI secretion ATPase
2.6	4148640-4148664	PA14_RS18920	AcrB/AcrD/AcrF family protein – transporter for acriflavin resistance
2.15	367555-367581	PA14_RS01670	amino acid ABC transporter permease
2.20	1770525-1770505	PA14_RS08260	ATP-dependent Clp protease ATP-binding subunit
1.2	6290248-6290234	PA14_RS28750	DNA-binding protein HU-alpha
2.17	6288448-6288430	PA14_RS28740	hypothetical protein
2.16	613388-613412	PA14_RS02845	reductase
2.6	2706841-2706859	PA14_RS12735	DNA-binding protein
2.18	484064-484082	PA14_RS02230	amine oxidase
2.8	6245496-6245512	PA14_RS28535	D-amino acid dehydrogenase small subunit
2.16	1580314-1580336	PA14_RS07385	alginate O-acetyltransferase
2.16	6100958-6100942	*cysQ*	3'(2'),5'-bisphosphate nucleotidase (sulfate assimilation, phosphatidylinositol phosphorylation)
2.16	6463649-6463631	PA14_RS29530	Ribonucleoside-diphosphate reductase, adenosylcobalamin-dependent

Table IV/2. Partial matches of CRISPR spacers with PA14 genome.

Spacer: [CRISPR locus].[spacer number]; Match: position of match in PA14 genome (NC_008463.1). Matches sorted from strongest to weakest.

a certain extent (Fig IV/5a), this was not as drastic at intermediate arabinose concentrations (0.1% (w/v)) as previously observed for Csy3::LacZ pHERD30T_SRM (Fig III/4), indicating that CRISPR-Cas may help to alleviate the cost of RM expression. Additionally, to determine whether unmethylated WT pHERD30T_SRM could adopt the same phenotype as premethylated WT pHERD30T_SRM in the presence of arabinose, we transferred WT SRM transformants into fresh medium containing 1% (w/v) arabinose daily. Each day, we evaluated whether the bacteria had acquired the resistance phenotype of WT pHERD30T_SRM transformants (Fig IV/1a) by performing plaque assays with DMS3vir. After three days, one of three replicates became entirely resistant to DMS3vir; both other replicates remained susceptible even after an additional transfer (Fig IV/5b). This result suggests that an RM system can become established without premethylation in WT bacteria when exposed to high arabinose concentrations. Unexpectedly, this occurs without a loss of RM function (which

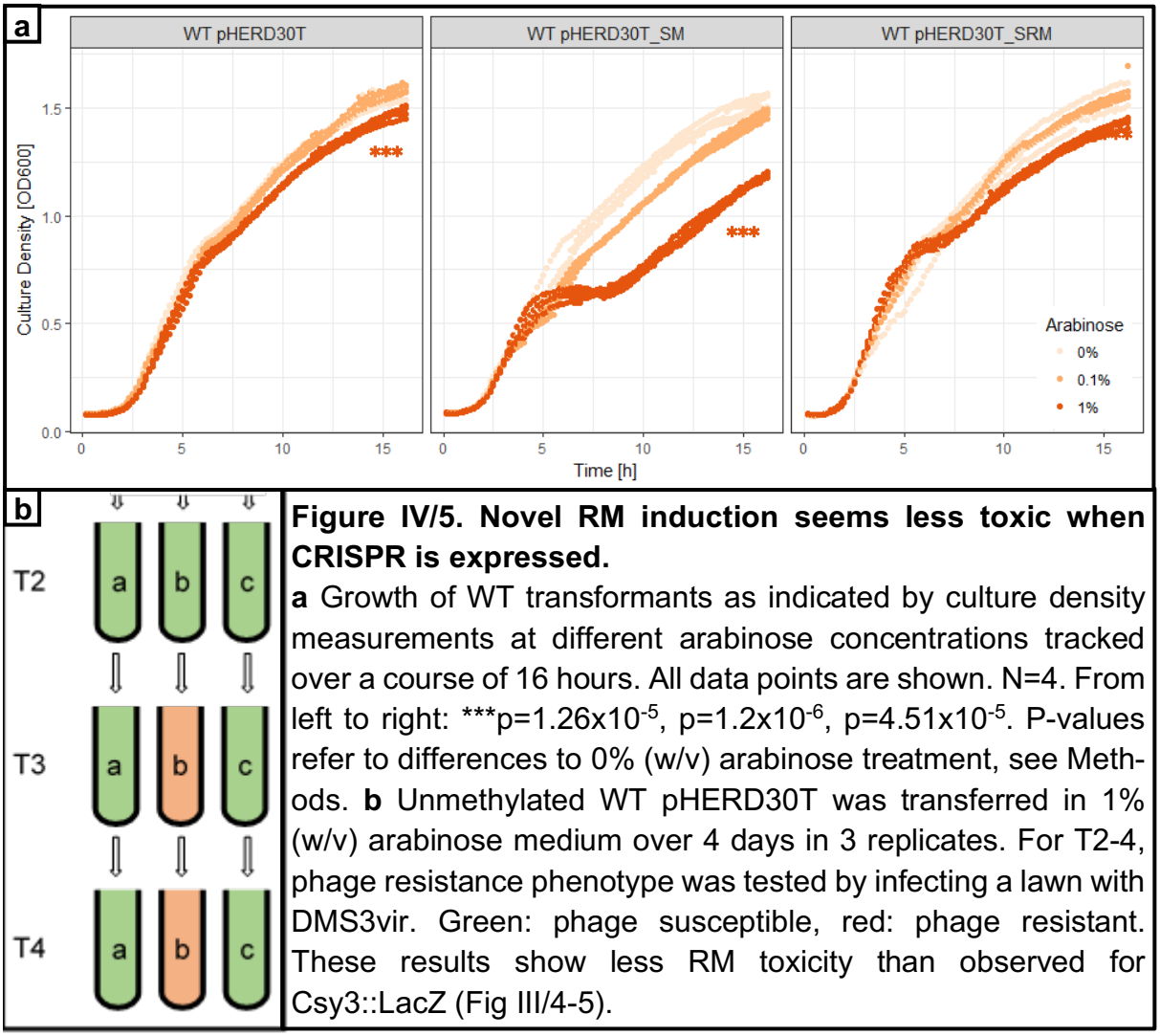


Figure IV/5. Novel RM induction seems less toxic when CRISPR is expressed.
a Growth of WT transformants as indicated by culture density measurements at different arabinose concentrations tracked over a course of 16 hours. All data points are shown. N=4. From left to right: ***p=1.26x10^{-5}, p=1.2x10^{-6}, p=4.51x10^{-5}. P-values refer to differences to 0% (w/v) arabinose treatment, see Methods. **b** Unmethylated WT pHERD30T was transferred in 1% (w/v) arabinose medium over 4 days in 3 replicates. For T2-4, phage resistance phenotype was tested by infecting a lawn with DMS3vir. Green: phage susceptible, red: phage resistant. These results show less RM toxicity than observed for Csy3::LacZ (Fig III/4-5).

we observed in Csy3::LacZ bacteria; Fig III/5) in some cases. Together, these

data suggest that toxicity of inducing a novel RM system is considerably lower when a functional CRISPR-Cas system is present, which is perhaps due to HsdR downregulation by a partial match between a CRISPR spacer and *hsdR*. For confirmation of this putative downregulation, it would be interesting to extract lysate from WT pHERD30T_SRM and compare it to lysate extracted from Csy3::LacZ pHERD30T_SRM strains in order to confirm whether HsdR is present at lower levels in WT transformants. Additionally, we suggest that more extensive analyses be carried out to determine whether toxicity is unique to Csy3::LacZ transformants or can be found in all strains lacking CRISPR.

Overall, our data leaves no doubt that there is a joint effect of CRISPR and RM on phage resistance in our model system. Most PA14 transformants expressing both CRISPR and RM show complete resistance against all pilus-specific phages and partial resistance against LPS-specfic LMA2, reducing its titres by 1-2 orders of magnitude. While all Cas proteins seem to be essential for this effect, resistance can be observed if HsdR is not expressed. This led us to believe that RM methylation patterns may lead to constitutive activation of CRISPR-*cas* loci, in turn altering downstream gene expression and ultimately leading to a lack of pilus expression. As aspects of our data remain unexplained by this model, future studies will be needed to shed light on the mechanistic basis of RM and CRISPR synergy in this model system.

Methods

Transformation of WT and CRISPR-knockout strains

Transformations were carried out as described in Chapter III. To generate premethylated RM-transformants, strains were initially transformed with pHERD20S_SM and grown in LB + 50 µg/ml streptomycin + 0.2% (w/v) arabinose over 3 daily transfers to ensure complete methylation of the bacterial genome. Afterwards, these strains were transformed with pHERD30T_SRM and henceforth grown in LB + GM50 + 0.2% (w/v) arabinose. As empty vector control, pHERD-30T was used for pHERD30T_SRM and pHERD30T_SM, and pCDF-1b for pHERD20S_SM constructs.

Resistance phenotype determination

Strains were grown overnight as indicated, then 300 - 600 µl of culture was mixed with 12.5 ml LB broth containing 0.5% (w/v) agar with 0.2% (w/v) arabinose and appropriate antibiotics, and plated onto a square plate of LB agar. Phage were diluted from 10^0 - 10^{-7} in M9 salts (0.6% $Na_2HPO_4 \cdot 7H_2O$, 0.3% KH_2PO_4, 0.05% NaCl, 0.1% NH_4Cl; all w/v) and 5 µl of each dilution were spotted onto bacterial lawns. After overnight incubation at 37°C, apparent titres could be calculated by counting plaques at appropriate dilutions. EOP was calculated as a measure of phage infectivity by (titre on pHERD30T_SRM strain) / (titre on empty vector control) as indicated. Similarly, for plaque assays 150 µl of overnight culture was mixed with 10 µl phage stock and 4 ml LB broth containing 0.5% (w/v) agar and antibiotics/arabinose as indicated; this was plated onto LB agar and incubated overnight at 37°C. Strains that showed levels of phage infection similar to empty vector control were scored as “susceptible”, and strains on which phage cannot plaque were scored as “resistance phenotype”.

Bacterial morphology

To investigate bacterial morphology, we streaked overnight cultures of different strains onto LB or LB + GM50 + 0.2% (w/v) arabinose and incubated these at 37°C for ~24 hours. Morphology was visually scored by similarity to controls as shown in Fig IV/3a.

Phage adsorption assays were carried out similar to a previously described protocol (Chibeu *et al.*, 2009). Overnight strains were diluted 1:100 in LB + 10 mM $MgSO_4$ + GM50 + 0.2% (w/v) arabinose and grown to mid-log phase for 2.5-3 h at 37°C. Pellets were gathered by centrifugation (15 min, 3500 rpm), washed in the same medium and resuspended to an OD600 of 2.0 to generate cultures of ~$3x10^7$ cfu/ml. 0.1 ml of these suspensions (or 0.1 ml medium for no-bacteria control) were mixed with 0.9 ml LB + 10 mM $MgSO_4$ + GM50 + 0.2% (w/v) arabinose containing ~$3x10^4$ pfu/ml of Φ1214 in three replicates, resulting in an MOI of 0.001. The mixes were incubated at 37°C for 8 minutes, after which cells were removed by centrifugation (14700 rpm for 5 min at 4°C), and 960 µl of supernatant was treated with chloroform (1:10, followed by centrifugation at 3500 rpm for 20 min at 4°C and filter-sterilisation of the supernatant) and spotted onto a lawn

of Csy3::LacZ in an M9 dilution series to determine the unbound phage titre. Percentage of bound phage was calculated as (3×10^4 - titre of unbound phage)/(3×10^4).

WT pHERD30T_SRM transfers, growth curves

Three transformants of WT pHERD30T_SRM were transferred into fresh medium (LB + GM50 + arabinose as indicated) daily. Samples of each timepoint, including T0, were frozen at -80°C in 20% (w/v) glycerol. At timepoints 2, 3, and 4, infectivity of ~ 2×10^7 pfu DMS3vir was tested on lawns containing 150 µl bacteria of each replicate. After an overnight incubation at 37°C, bacterial phenotype was visually determined: plates with a smooth bacterial lawn without plaques indicated a resistance phenotype, while normal phenotype plates showed hardly any bacterial growth.

Growth of WT transformants was measured as described in chapter III.

WT pHERD30T_SRM evolution

Six pHERD30T_SRM transformants (A-F) and WT pHERD-30T were incubated together with DMS3vir, JBD5 (MOI ~20) or LB medium (in an equal volume to phage added as a no-phage control) at 37°C, and 50 µl of these cultures were transferred into fresh medium (LB + GM50 + 0.2% (w/v) arabinose) daily. Additional controls included Csy3::LacZ pHERD30T_SRM and Csy3::LacZ pHERD30T. With each transfer, an additional 10^9 pfu of appropriate phage or LB medium were added to each replicate. Samples of each timepoint, including T0, were frozen at -80°C in 20% (w/v) glycerol. 50 µl of a 10^{-5} dilution of T3 samples were plated onto LB + GM50 + 0.2% (w/v) arabinose and incubated overnight at 37°C.

To determine whether CRISPR spacer acquisition had occurred, 21 individual colonies per replicate were suspended in 10 µl H_20 each, and 2 µl of resuspensions used as a sample for a colony PCR. Primers 7 and 8 were used to amplify CRISPR1, and primers 10 and 11 to amplify CRISPR2 (Table S1). The PCR was carried out in a 10 µl reaction volume using Thermo Scientific's DreamTaq PCR Master Mix according to the manufacturer's instructions (T_{melt} = 50°C, $t_{elongation}$ = 45 sec). 5 µl of amplifications were spotted onto a 2% (w/v) agarose gel and separated by electrophoresis for visualisation. Colonies that acquired a spacer had integrated it into the CRISPR array, resulting in a larger PCR fragment.

Partial spacer matches in PA14 genome

We used CRISPRfinder (Grissa, Vergnaud and Pourcel, 2007a) to identify all 35 spacers present in PA14 WT (14 in CRISPR1, 21 in CRISPR2). Partial spacer matches to the complete PA14 genome (NC_008463.1) were identified using CRISPRTarget (Biswas *et al.*, 2013), disregarding the exact matches of the spacers with themselves in the CRISPR loci. Restrictions included a seed region at the 3' end of the protospacer (exact matches at nts 1-8, except nt 6) and, where mentioned, a PAM (protospacer adjacent motif) of GG adjacent to the 3' end. Results were expanded to include all partial matches without a mismatch cutoff. The genes partial matches are found in were identified as listed on Genbank (Clark *et al.*, 2016).

Statistical Analyses

Statistical analyses were carried out using R software, specific tests used for each experiment are highlighted in figure legends. For bacterial growth curves (Fig IV/5), the package Growthcurver (Sprouffske and Wagner, 2016) was used to estimate various growth curve statistics. As a proxy for bacterial growth, carrying capacity k of bacterial growth at moderate (0.1% w/v) and high (1% w/v) arabinose concentrations for each strain was compared to k for its respective growth in the absence of arabinose and their difference assessed with an analysis of variances and a Tukey's post-hoc test.

General Discussion

Diversity-generating mechanisms (DGMs) provide a large benefit to hosts and can protect populations from parasite epidemics, accordingly species across the tree of life have evolved such mechanisms. CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repats-CRISPR associated) is traditionally thought of as the DGM in bacteria, but due to the nature of its HsdS specificity subunit, some Type I Restriction-Modification systems have the capacity to generate diversity too. In *Mycoplasma pulmonis*, the Type I RM Mpu shufflon generates novel specificities by recombining its *hsdS* pseudogenes. In this way it can generate 30 unique sequences, 14 of which are most likely to be biologically active as small and standard sized subunits with one or two target recognition domains (TRDs) respectively. While we generated sequences of all possible *hsdS* conformations and some protein models, the functionality of small and large sized subunits remains to be investigated *in vivo*.

When expressed in CRISPR-deficient *Pseudomonas aeruginosa* PA14, this Type I RM system is toxic. This autoimmunity can be alleviated by pre-methylating the bacterial genome before introducing the full RM system, but RM only provides low-level immunity against some *Pseudomonas* phages. This shows that RM may have more associated costs than benefits in this system, and further raises the question of whether shufflons are associated with the same autoimmune costs in nature. To determine whether overexpression of this RM system is generally costly, future studies could involve designing a RM construct with weaker ribosome binding sites to trial the impact of decreased RM protein production on toxicity. In order to confirm whether a shufflon specificity switch is associated with the same costs as introducing an entirely novel RM system, future studies should simulate such a switch. This would require encoding the SRM construct with a different *hsdS* subunit on a different vector backbone and transforming a strain with an established RM with this new construct. If toxicity costs are transferable to this specificity switch scenario, it postulates that certain *hsdS* conformations may be very stable as additional inversions are highly unfavourable due to autoimmune costs and will therefore only occur under high phage selective pressures. Investigating natural populations of *M. pulmonis* and other shufflon-encoding bacteria for the abundance of RM specificities over time could give an indication of whether such stable *hsdS* conformations exist.

When CRISPR-Cas and RM are coexpressed, PA14 becomes completely resistant to most *Pseudomonas* phages tested and partially resistant to LPS-specific LMA2. Previous studies had indicated that RM could aid CRISPR-Cas by feeding RM-degraded DNA fragments as substrates to enhance spacer acquisition (Dupuis *et al.*, 2013; Hynes, Villion and Moineau, 2014). While our data does show resistance and even spacer acquisition when CRISPR and RM are jointly expressed, other observations contradict the theory that the RM system is stimulating CRISPR by creating spacer substrates. Most strikingly, resistance can still be observed when restriction endonuclease HsdR is not present. Additionally, different levels of resistance to pilus or LPS-specific phages remain unexplained by this model. Therefore, resistance must be due to other interactions between CRISPR-Cas and RM. As our data conflicts with several explanations for the synergistic phage resistance when CRISPR and RM are coexpressed (such as CRISPR-dependent pilus downregulation or other interactions between RM and CRISPR), future experiments need to address the fundamental flaws in these models.

Gene regulation

With a partial match between the spacer and target, CRISPR-Cas can show gene regulatory rather than nucleolytic activity (Westra, Buckling and Fineran, 2014). This has been exploited in synthetic biology, where nucleolytically deficient Cas9 variants are used to up- or downregulate gene expression (Lo and Qi, 2017). Therefore, we proposed a model in the previous chapter in which CRISPR-Cas and RM provide phage resistance by causing downregulation of the pilus gene. This would require RM methylation patterns to constitutively activate CRISPR-Cas and in turn CRISPR-Cas to downregulate pilus expression. While this model explains how resistance can occur in the absence of HsdR, we would expect ΔCas1 (this nuclease is only important in spacer acquisition) pHERD30T_SRM transformants to also be resistant (unless Cas1 deletion impacts Csy gene transcription or translation). This indicates that, if CRISPR-mediated gene regulation is occurring, it is probably through a different mechanism than previously described for other CRISPR systems (e.g. Type II systems, which utilise non-CRISPR sca-RNA to modulate expression of BLP-1 (Sampson *et al.*, 2013)).

To gather more evidence in support of gene regulation occurring, future studies need to identify the MpuUI RM recognition site. This can be done by analysing

methylomes of RM transformants and comparing them to empty vector strains. Alternatively, to more directly test whether RM activity enhances CRISPR expression, a Miller assay could measure β-galactosidase levels (Schaefer *et al.*, 2016) in Csy3::LacZ pHERD30T_SRM transformants. In Csy3::LacZ, *lacZ* replaces *csy3*, so β-galactosidase levels will be a direct representation of CRISPR activity. Therefore, if RM methylation enhances CRISPR activity, β-galactosidase will be higher in Csy3::LacZ pHERD30T_SRM than in empty vector controls. Additionally, other direct tests could confirm whether WT pHERD30T_SRM bacteria are surface mutants that have lost or downregulated their pilus. To these means, we could carry out motility or competition assays with empty vector controls (bacteria without their pilus are generally less fit than WT bacteria). Furthermore, we could extract protein lysate from WT pHERD30T_SRM and control bacteria and blot it with an antibody specific to the pilus to confirm whether WT pHERD30T_SRM expresses it at lower levels than WT bacteria. As such an antibody is not commercially available, this experiment would involve isolating pilus proteins and producing polyclonal antibodies against them. Alternatively, a qPCR assay could indicate whether pilus expression is lower in RM transformants. Finally, it should be tested whether knocking out Cas1 impacts Csy protein production to further validate this hypothesis.

General DNA damage response

Perhaps the increased phage resistance observed when RM and CRISPR are coexpressed can be attributed to a more generalised response to DNA damage, which could be induced when CRISPR targets the bacteria's genome with a partial spacer mismatch. In a series of previous studies (Zegans *et al.*, 2009; Cady and O'Toole, 2011; Heussler *et al.*, 2015) an inhibition of *P. aeruginosa* biofilm formation was attributed to CRISPR-Cas targeting lysogenised DMS3 with a partial spacer match, which resulted in damaged rather than cleaved DNA. This DNA damage induced the SOS-response, which was toxic to bacteria in biofilms in presence of phage protein. Therefore, even if RM methylation constitutively activates CRISPR-Cas as in the gene regulation model above, this may not lead to direct regulation of gene expression by CRISPR, but rather induction of the SOS or a generalised DNA damage response that causes pilus downregulation. In future, this could be tested by experimentally inducing DNA damage or the SOS-

response in PA14 WT, and determining whether this leads to pilus downregulation.

Protein-protein interaction

As an alternative model, phage resistance may be due to interactions between RM and Cas protein complexes that promote spacer acquisition – not, as proposed by Hynes, Villion and Moineau (2014), RM fragments acting as templates for spacer acquisition. This protein-protein interaction would have to occur between the M_2S_1 complex and CRISPR-Cas, perhaps with the RM specificity subunit leading CRISPR spacer acquisition machinery towards phage DNA. One piece of evidence that points towards this model rather than gene regulation is spacer acquisition after three days of evolution with DMS3vir. Typically, populations that evolve CRISPR-resistance acquire various, often multiple spacers (Morley *et al.*, 2017), however in our experiment no clone acquired more than one spacer. If future sequencing of these spacers should reveal that they are all identical, it would indicate that spacer acquisition was a rare event and only occurred once throughout the evolution experiment. Were this the case, this spacer would be associated with a clear fitness benefit (as the initial clone carrying it expanded throughout half the population), but it would diminish the role of RM enhancing spacer acquisition (as it would be a rarer event).

Overall, it seems that the gene regulation model, albeit by an unexplained mechanism, is the most likely of these alternative explanations. Bearing this in mind, we should trial the SRM construct with alternative *hsdS* conformations, which are predicted to have a different sequence specificity. These might not have such a drastic impact on CRISPR gene regulation due to different methylation patterns, and may allow us to investigate RM and CRISPR interactions without influencing pilus expression.

In this thesis, I reviewed the literature to argue that different types of DGMs found across the entire tree of life lead to distinct coevolutionary dynamics depending on the level of diversity they generate. Specifically, the targeted bacterial DGMs CRISPR-Cas and diversity-generating Type I RM systems are crucial for bacteria-phage coevolution. I analysed the *M. pulmonis* Mpu shufflon *in silico* and found that it has the capacity to generate 30 different specificity subunits, 12-14

of which are most likely to be biologically active due to their conventional number of 1-2 TRDs. I adapted this shufflon for expression in *P. aeruginosa* PA14 to create an experimental system of studying bacteria-phage interactions in a population with diverse RM specificities, and also to examine the joint impact of CRISPR-Cas and RM. Initially, I found that induction of a novel RM system is toxic while only providing low-level resistance in this model system, which indicates shufflons may have larger costs and smaller benefits than anticipated. Finally, I coexpressed CRISPR-Cas and RM in a PA14 host and found unprecedented levels of complete resistance against most phages, which may be due to CRISPR and RM-dependent pilus downregulation. The true mechanism of complete and partial resistance in RM and CRISPR coexpression remains to be revealed.

Appendix

Glossary

Monoculture effect The increased incidence of diseases in monocultures of the same crop.

Indels The insertion or deletion of bases in the DNA of an organism.

Transposable element A DNA sequence that can mobilize to a new position within the genome.

Genetic drift A change in allele frequencies as a result of the random sampling of gametes that form the next generation.

Parthenogenic Reproducing in an asexual manner.

Germinal centres Sites within secondary lymphoid tissue where B cell proliferation, selection and maturation take place during antibody responses.

CRISPR escape phage Phage that acquire mutations at positions in the protospacer (the sequence matching the CRISPR spacer) or the protospacer adjacent motif (a short DNA sequence required for CRISPR activity) that allow them to overcome CRISPR-Cas immunity.

Arms-race dynamics (ARD). Co-evolutionary dynamics that are characterized by the increase of both host resistance and pathogen infectivity ranges: hosts evolve resistance to a broader range of pathogen genotypes and pathogens evolve infectivity to a broader range of host genotypes.

Fluctuating selection dynamics (FSD). Co-evolutionary dynamics that are characterized by fluctuations in host and pathogen genotypes owing to frequency dependent selection, whereby the fitness of host genotypes is inversely correlated with their frequency in the population.

Tables

Plasmids and Vectors			
Plasmid	**Reference**	**Antibiotic Resistance gene**	**Restriction Enzymes used for cloning**
pHERD-30T	(Qiu *et al.*, 2008)	Gentamycin	n/a
pHERD-20T	(Qiu *et al.*, 2008)	Ampicillin	n/a
pCDF-1b	Novagen vector	Streptomycin	n/a
SRM construct	This study	n/a	n/a
pHERD30T_SRM	This study	Gentamycin	NcoI, BamHI (subcloning) NcoI, HindIII
pHERD30T_SM	This study	Gentamycin	KpnI
pHERD20S_SM	This study	Ampicillin, Streptomycin	NcoI, HindIII (SM genes) AgeI, SgrAI (Strep resistance)
Primers			
Primer	**Sequence (5' → 3')**		**Usage**
pBAD forward	ATGCCATAGCATTTTTATCC		pBAD forward or Bam/AraI & R reverse amplify *hsdS*
R reverse	AGCAGTTCGTTGCGGGACAT		
Bam/AraI	CAAAGCCATGACAAAAACGC		
S-KpnI-bwd	AAGGTACCCTATTCGTCCTTGATCTTTTC		Amplify pHERD30T_SRM while deleting *hsdR*, insert KpnI restriction sites at either end.
M-KpnI-fwd	CCGGTACCGAATCCAAGATCTAAAGT		
7	CTAAGCCTTGTACGAAGTCTC		Amplify PA14 *CRISPR1*
8	CGCCGAAGGCCAGCGCGCCGGTG		
10	GCCGTCCAGAAGTCACCACCCG		Amplify PA14 *CRISPR2*
11	TCAGCAAGTTACGAGACCTCG		

Table S1: Plasmids, Vectors, and Primers used throughout the thesis.

Bacterial Strains		
Strain	**Shorthand name**	**Description/Reference**
NEB 5-alpha competent *Escherichia coli fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	*E. coli*	Used for cloning. Commercially available from New England Biolabs.
Pseudomonas aeruginosa UCBPP-PA14	WT	
UCBPP-PA14 *csy3::lacZ*	Csy3::LacZ	(Cady *et al.*, 2012)
UCBPP-PA14 *ΔCRISPR, ΔCas*	ΔCR/Cas	(Cady and O'Toole, 2011)
UCBPP-PA14 *ΔCRISPR1, ΔCRISPR2*	ΔCR1,2	(Cady and O'Toole, 2011)
UCBPP-PA14 *Δcas1*	ΔCas1	(Cady and O'Toole, 2011)
UCBPP-PA14 *Δcas3*	ΔCas3	(Cady and O'Toole, 2011)
UCBPP-PA14 *Δcsy1*	ΔCsy1	(Cady and O'Toole, 2011)
UCBPP-PA14 *Δcsy2*	ΔCsy2	(Cady and O'Toole, 2011)
UCBPP-PA14 *Δcsy4*	ΔCsy4	(Zegans *et al.*, 2009)

Bacteriophages		
Phage	**Reference**	**Genome size [nt]**
DMS3vir	(Cady and O'Toole, 2011)	36415
Φ1214	(Lindberg and Latta, 1974)	37053
JBD5	(Bondy-Denomy *et al.*, 2013)	37740
LUZ24	(Ceyssens *et al.*, 2009)	45625
Φ68	(Lindberg and Latta, 1974)	48097
JBD18	(Cady *et al.*, 2012)	39014
JBD25	(Cady *et al.*, 2012)	39552

Table S2: Bacterial Strains and phage used throughout the thesis.

Construct sequences

Sequences in 5'-3' direction of all synthetic Mpu shufflon sequences adapted for expression in *Pseudomonas aeruginosa* PA14 are listed below.

Yellow: Restriction sites
Green: T7 Ribosome Binding Site
Blue: Codon optimised for *Pseudomonas aeruginosa* PA01 for restriction site avoidance

>SRM construct
CCATGGAAATCTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGGGCAAGTCCAAGTA
CAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTACTCCTCCAAGACC
AAGGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAACGGCGAATACATCC
TGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCAACGAAAAGTTCTC
CACCACCTCCAACTGCTTCATCCTGAAGGTCAACGAAAACATCGTCAAGACCAAGTTC
CTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATGGCCATCGGCTCCG
CCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCAACCTGCCGAACCT
GAAGATCCAGTCCGCCATCATCAAGATCATCGAACCGAAGGAAGACCTGTTCTTCCGC
CACAAGAACCTGGTCCGCATCGACTCCGAAGAAAACACCAAGAAGGACCTGTCCATCC
TGATCAAGATCATCGAACCGCTGGAAAAGCAGATCAACGCCTTCGACGAACTGATCCT
GTCCGAACAGAAGTCCCTGCAGCACTACCTGAACTACTTCCTGAACAAGCTGGCCTCC
ATCAACCCGTCCATCTTCAAGAACTACAAGCTGGGCGAAATCGCCAAGATCCTGTCCG
GCAAGACCCCGTCCACCGCCAAGAAGGAACTGTGGAAGAAGGAAATCCCGTTCTTCGG
CCCGGGCGACCTGGACAACATGGTCCCGAAGCGCTTCATCACCTTCAACGAAAAGATG
ATCAAGCGCTCCGGCACCATCCTGTTCTCCTCCGCCGCCACCATCGGCAAGGTCGGCA
TCCTGGACAACCTGTCCTGGTTCAACCAGCAGATCACCTCCATCGAAGCCAACAACAA
CTACGTCATGGACAAGTTCCTGTTCTTCCTGCTGAAGAAGATCTCCTCCAAGATCAAG
TTCGAAAACTCCTCCGGCACCATCTTCCCGACCATCAAGAAGAAGTACTTCGAAAACT
TCACCCTGGAAATCCCGAACCTGAAGACCCAGTCCGCCATCCTGGGCATCATCGAACC
GCTGCACAAGAAGATCAACCTGCTGAAGCAGAAGAAGAAGCTGCTGGAAAAGCGCTTC
ATCTACTACCAGAACCACCTGATCAAGGAAAAGATCAAGGACGAATAGGGTACCTTAA
CTTTAAGAAGGAGATATACATACCCATGTCCCGCAACGAACTGCTGTACGAAAAGGAA
TTCGTCGACGACCTGGTCAAGAACCAGAAGTACGTCAAGCTGGACATCAAGAACGAAG
AAAAGATCTTCGAACTGATCTTCGAAAACATCGGCCGCCTGAACAACATCGAACTGAC
CCAGCAGAACATCCTGGACATCCGCCGCGAACTGCTGTCCAACAACTCCGCCTCCTCC
TACCGCTTCGCCCAGTACCTGTGGGGCTTCGACACCGTCAAGATCTACAAGAACGACG
GCCCGAAGAAGATCCGCCTGAAGTTCGTCGACTGGGAAAACTGGGCCAACAACGAATT
CTACGTCCTGCAGCAGTTCCCGACCCGCGACGCCAAGAACAACATGGGCAAGCGCTTC
GACTCCCTGATCCTGATCAACGGCTTCCCGCTGATCCTGTTCGAATTCAAGGACAAGT
CCGAAAACATCAACAAGGCCATCAACCAGATCGACGAATCCTACCGCGGCTCCGTCCT
GAACAAGGGCATCTTCCGCTTCATCCAGATCCTGATCGGCTCCAACTTCGAAGAAGTC
AAGTTCCTGGCCAACAACAAGCGCACCAACAACAACAAGATCCTGTCCTTCAAGTGGA
CCTCCGAAAACGGCGGCTCCTCCAAGGAACTGATCAAGGACTTCCTGAACAAGAACGC
CCTGGAAGAATACCTGAAGAACTACGTCGTCTTCCAGCGCTCCAAGGACGACGAAAAG
ATCATCCTGCTGCGCCCGTACCAGCAGCGCGCCATCAAGAAGGCCATCAACTTCGTCG
AAAAGCAGCTGAAGACCAACCTGGACGCCAAGCACAACCTGAACAACGCCTACATCTG
GCACACCACCGGCTCCGGCAAGACCCTGACCTCCTACAAGATCGCCGAAATCCTGTCC
AAGAACTCCGACATCGACCACGTCGTCTTCCTGGTCGACCGCAACGACCTGAACGACC
AGACCTCCCAGACCTTCCAGAAGCTGATGTCCTCCTCCAAGAACGAAAAGATCGACTT
CCTGAACCAGGACACCTCCAAGGACCTGTACGAAATCTTCCTGAAGAAGGAAAAGCTG
ATCATCACCACCATCCAGAAGCTGAACAACATCCTGTCCTCCTACAAGAACGAAAAGA
TCGAATTCCTGACCAACAAGAAGTTCGTCTTCATCATCGACGAATGCCACCGCTCCAA
CGCCGGCCTGATGGGCAAGCGCATCAAGGACTTCCTGAACAACTCCATCATGATCGGC
TTCTCCGGCACCCCGATCTTCGAAGAAAACAACGACCGCGAAACCCAGAAGATCTTCG
GCAACGAAATCGACTCCTACAACATGAAGGACGCCATCCTGGACAAGAACGTCCTGGG
CTTCAAGGTCGTCAACTACTACCAGGAAACCCGCATCTTCCGCGAAAACAACAACTCC
AACCTGGGCAAGATCAAGTCCATCATCAACGTCATCAAGTCCAAGCACCTGGACTTCA
CCAACAACCGCAACTACAACTCCATCATCGCCTTCGACACCATCCAGGACGCCCTGAC
CTTCTACGACGAATTCTACAAGATGGACGTCGGCGACATCTTCGCCACCCCGATCTTC

TCCTCCTACTCCAACGAAGAAAAGAACGAAAAGTTCTTCAACCTGAAGGAACACAAGG
AAAAGATCCTGAAGCGCTACGAAGAAAAGTTCAACACCTCCTTCAAGGTCGAAGACTT
CGACAAGTACGTCAACGACGTCCAGTGGCGCTTCAAGGAATACAACTCCGAAAACAAC
TCCATCGACATCGTCATCGTCGTCGACATGCTGCTGACCGGCTTCGACTCCCCGCGCA
CCAACACCCTGTACATCAACAAGGAACTGAAGAACCACAACCTGATCCAGGCCTTCTC
CCGCACCAACCGCCTGTCCGACTACTCCAAGAAGCGCGGCATCATCGTCAACTTCTCC
CTGGAAGAACAGTCCATCAACGACGCCTTCAAGATCTACGCCAACTCCTCCGACAAGG
AAATCCAGCAGCTGGTCTACGGCGAAAAGTACGAACAGGTCGTCGAAGACTTCATCAA
CTTCTGGAACTCCCTGAAGATCTCCTTCTCCAACATCTACGACGAAAAGAACAACGAA
ATCTTCCGCAACATCTCCCTGGAAAACAAGAAGAAGTACCTGAAGAACCTGTCCCAGG
TCTCCAACATCTTCTCCTCCCTGAAGACCTTCAAGGAATACGGCAAGAACGAAAAGAT
CTCCGACTTCTCCCTGGAACAGCTGAACCAGTACCAGAAGTGGGCCAACGAAATCAAG
AAGAACCTGTCCACCAACGAAAAGGAAAAGATCTCCTACGAAGTCCTGAACTCCATCG
ACATCTCCAACATCAAGTTCGCCTACAAGGAAATGATCATCGACGAAATCTACCTGGA
AAACCTGCTGTTCTTCAACAAGAAGATCTCCAAGTACCCGAACAACCGCCTGACCTAC
GAAGACACCCTGTCCGAAATCGACAAGCACATCCAGCTGATCAAGAACAACTACAACC
AGGGCAAGATCAACCAGAAGGAATACGAAATCTTCCTGCTGCTGGTCCAGAAGTGGAA
GAACGAAATCAAGAACTTCTTCATCAAGAAGGACAAGTCCCTGGACGAAAAGGAATTC
ATCGACTACGGCAAGCGCATCCTGAAGTCCGTCTTCCAGAAGGTCAAGAACCAGATCG
AAGCCATGTGGCTGGAAAAGATCCTGAAGGAATACCACGGCATCAACAACGACCAGAT
CCGCAAGGACTGGAAGAAGCGCATCAACGACAAGGACCTGGACGACATCGAAAAGTCC
GAATTCATCAAGAAGTGGTCCCGCCGCTCCAAGGAAGTCGACAAGGACATCATCGACA
AGCTGTCCATCGAATACAAGGAATCCATCGAAGCCTTCCTGGACTTCGAAATCAAGAT
GAACAAGATCATCGAATCCAAGATCTAAAGTTTAACTTTAAGAAGGAGATATACATAC
CCATGTCCAACTCCAAGGAACTGATCGCCGTCGTCAAGAAGATCTGCGACCAGCTGCG
CTCCAAGATGGAAGTCACCGAATACCGCGACTACATCATGGGCTTCCTGTTCTTCAAG
TACCTGTCCGAACAGTCCGAAAAGAACTTCGAAGAATTCAAGGAACGCGTCGACTACA
TCAAGTACTCCGAATTCGACGAAAACCACGAACAGTTCAAGAAGATCAAGGAAATCAT
CATCCAGAACGACGACGACTTCTTCCTGGCCTACAAGTACTCCTTCCAGAACGTCGTC
GACATGATGAACCAGGGCAAGAACGTCATCCCGACCATCGAAGAATCCTTCAACAAGA
TCGAATCCATCAACTCCGAACTGAACGACGAAAAGAAGGAATTCTTCAAGGACCTGTT
CACCAACATCGACTTCTCCAACAAGAACCTGGGCAACATCGACGAAGAAAAGGAAAAG
ACCATCCAGCTGATCATCAAGGAAATCAACACCCTGAACCTGAGCATGGACGAAGTCG
ACCACTTCGGCAACACCTACGAATACCTGCTGTCCGAATTCGCCTCCGACACCGGCAA
GAAGGCCGGCGAATTCTACACCCCGTCCAAGGTCTCCGAACTGCTGGTCAAGATCGTC
TCCCACGGCAAGAACAAGATCAACAAGGCCTACGACCCGGCCTGCGGCTCCGGCTCCC
TGCTGATCAAGCTGGCCAACAAGGTCGGCAAGTACAACAAGATCTACGGCCAGGAAGT
CAAGACCGCCACCTACAACCTGGCCCGCATGAACTTCATCCTGCGCGGCGTCCCGTTC
TCCAAGCTGGACCTGCGCTCCGGCGACACCCTGATCAACCCGCTGCACATCGAAGAAG
AAGGCTCCTTCGACTGCATCGTCGCCAACCCGCCGTTCTCCCAGAAGTGGAACCCGAC
CCAGGAACTGTCCAAGGACCGCCGCTACAACTCCTACCCGTCCCTGGCCCCGAAGTCC
TACGCCGACTTCGCCTTCCTGCAGCACATGCTGTTCCACGTCAACAAGGACAACGGCA
TCATCGCCTCCGTCTTCTCCCTGGGCATCCTGTCCCGCAAGTCCCCGAAGGCCGAAGA
AGACATCCGCAAGTACATCATCGACAAGAACTACATCGACACCATCATCTTCCTGCCG
CCGAACCTGTTCTACAACACCTCCATCGAATCCTGCATCATCGTCGCCCGCAAGAACA
AGCCGACCAACGACAAGCGCATCTTCATGATCAACGCCACCAAGGAATTCCAGAACGC
CAAGAAGCAGAACACCCTGTCCGACGAAAACATCAACCGCATCTTCTCCGCCTGGAAG
GAAAAGCGCGAAGAAGAAAACTTCTCCAAGTACATCTCCTACGAAGACATCGTCAAGA
ACGAATACTCCCTGTCCATGCGCTTCTACGACCTGGACAACTTCGACGAAGAATCCGA
AGACATCGACATCGACTTCGTCGAATCCGAAATCGTCAAGATCAACGAAGAACTGCTG
AAGTACGAAAACGAATTCAAGAAGAACCTGAACGAATTCCTGAACAAGAAGAACTAAG
GATCC

>MpuUII
CCATGGAAATCTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGGGCAAGTCCAAGTA
CAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTACTCCTCCAAGACC
AAGGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAACGGCGAATACATCC
TGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCAACGAAAAGTTCTC
CACCACCTCCAACTGCTTCATCCTGAAGGTCGACGAAAACATCGCCAAGACCAAGTTC
CTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATGGCCATCGGCTCCG

CCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCAACCTGCCGAACCT
GAAGACCCAGTCCGCCATCATCAAGATCATCGAACCGAAGGAAGACCTGTTCTTCCGC
CACAAGAACCTGGTCCGCATCGACTCCGAAGAAAACACCAAGAAGGACCTGTCCATCC
TGATCAAGATCATCGAACCGCTGGAAAAGCAGATCAACGCCTTCGACGAACTGATCCT
GTCCGAACAGAAGTCCCTGCAGCACTACCTGAACTACTTCTTCGGCAAGTTCTACCAG
ATCGAACCGTCCCTGTTCCACGACTACAAGCTGGAAAAGATCGCCAAGATCCGCCGCG
GCAAGATCATCAACTCCTTCGACCTGAAGGAAAACCCGGGCGACTACCCGGTCATCTC
CTCCAACACCAAGAACAACGGCATCTTCGGCTACCTGAACTCCTACATGTACGACGGC
GAATACATCACCATCTCCGCCGACGGCGCCTACGCCGGCACCGTCTTCCTGAACAACG
GCAAGTTCTCCATCACCAACGTCTGCTTCATCCTGCTGCTGAACGACAAGGTCAACCT
GCTGACCAAGTTCCTGTTCTACTACCTGAAGAAGAACGAAAACATCATCCAGAAGAAG
TCCATCGTCGGCTCCTCCCGCCCGTCCGTCCGCGAATACACCCTGTCCGAAATCGCCA
TCAAGATCCCGTCCCTGGAAATCCAGTCCGCCATCCTGGGCATCAACGAACACTTCCA
CTACGTCTAAGGTACC

>MpuUIII
CCATGGAAATCTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGGGCAAGTCCAAGTA
CAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTACTCCTCCAAGACC
AAGGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAACGGCGAATACATCC
TGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCAACGAAAAGTTCTC
CACCACCTCCAACTGCTTCATCCTGAAGGTCAACGAAAACATCGTCAAGACCAAGTTC
CTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATGGCCATCGGCTCCG
CCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCAACCTGCCGAACCT
GAAGATCCAGTCCGCCATCATCAAGATCATCGAACCGCTGGAAAAGCAGATCAACGCC
TTCGACGAACTGATCCTGTCCGAACAGAAGTCCCTGCAGCACTACCTGAACTACTTCT
TCGGCAAGTTCTACCAGATCGAACCGTCCCTGTTCCACGACTACAAGCTGGAAAAGAT
CGCCAAGATCCGCCGCGGCAAGATCATCAACTCCTTCGACCTGAAGGAAAACCCGGGC
GACTACCCGGTCATCTCCTCCAACACCAAGAACAACGGCATCTTCGGCTACCTGAACT
CCTACATGTACGACGGCGAATACATCACCATCTCCGCCGACGGCGCCTACGCCGGCAC
CGTCTTCCTGAACAACGGCAAGTTCTCCATCACCAACGTCTGCTTCATCCTGCTGCTG
AACGACAAGGTCAACCTGCTGACCAAGTTCCTGTTCTACTACCTGAAGAAGAACGAAA
ACATCATCCAGAAGAAGTCCATCGTCGGCTCCTCCCGCCCGTCCGTCCGCGAATACAC
CCTGTCCGAAATCGCCATCAAGATCCCGTCCCTGGAAATCCAGTCCGCCATCCTGGGC
ATCATCGAACCGCTGCACAAGAAGATCAACCTGCTGAAGCAGAAGAAGAAGCTGCTGG
AAAAGCGCTTCATCTACTACCAGAACCACCTGATCAAGGAAAAGATCAAGGACGAA-
TAGggtacc

>MpuUIV
CCATGGAAATCTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGGGCAAGTCCAAGTA
CAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTACTCCTCCAAGACC
AAGGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAACGGCGAATACATCC
TGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCAACGAAAAGTTCTC
CACCACCTCCAACTGCTTCATCCTGAAGGTCGACGAAAACATCGCCAAGACCAAGTTC
CTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATGGCCATCGGCTCCG
CCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCAACCTGCCGAACCT
GAAGACCCAGTCCGCCATCATCAAGATCATCGAACCGAAGGAAGACCTGTTCTTCCGC
CACAAGAACCTGGTCCGCATCGACTCCGAAGAAAACACCAAGAAGGACCTGTCCATCC
TGATCAAGATCATCGAACCGCTGGAAAAGCAGATCAACGCCTTCGACGAACTGATCCT
GTCCGAACAGAAGTCCCTGCAGCACTACCTGAACTACTTCCTGAACAAGCTGGCCTCC
ATCAACCCGTCCATCTTCAAGAACTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGG
GCAAGTCCAAGTACAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTA
CTCCTCCAAGACCCGCGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAAC
GGCGAATACATCCTGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCA
ACGAAAAGTTCTCCACCACCTCCAACTGCTTCATCCTGAAGGTCAACGAAAACATCGT
CAAGACCAAGTTCCTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATG
GCCATCGGCTCCGCCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCA
ACCTGCCGAACCTGAAGATCCAGTCCGCCATCCTGGGCATCATCGAACCGCTGCACAA
GAAGATCAACCTGCTGAAGCAGAAGAAGAAGCTGCTGGAAAAGCGCTCCATCTACTGC
CAGAACCACCTGATCAAGGAAAAGATCAAGGACGAATAGGGTACC

>MpuUV
CCATGGAAATCTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGGGCAAGTCCAAGTA
CAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTACTCCTCCAAGACC
AAGGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAACGGCGAATACATCC
TGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCAACGAAAAGTTCTC
CACCACCTCCAACTGCTTCATCCTGAAGGTCAACGAAAACATCGTCAAGACCAAGTTC
CTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATGGCCATCGGCTCCG
CCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCAACCTGCCGAACCT
GAAGATCCAGTCCGCCATCATCAAGATCATCGAACCGCTGGAAAAGCAGATCAACGCC
TTCGACGAACTGATCCTGTCCGAACAGAAGTCCCTGCAGCACTACCTGAACTACTTCT
TCGGCAAGTTCTACCAGATCGAACCGTCCCTGTTCCACGACTACAAGCTGGAAAAGAT
CGCCAAGATCCGCCGCGGCAAGATCATCAACTCCTTCGACCTGAAGGAAAACCCGGGC
GACTACCCGGTCATCTCCTCCAACACCAAGAACAACGGCATCTTCGGCTACCTGAACT
CCTACATGTACGACGGCGAATACATCACCATCTCCGCCGACGGCGCCTACGCCGGCAC
CGTCTTCCTGAACAACGGCAAGTTCTCCATCACCAACGTCTGCTTCATCCTGCTGCTG
AACGACAAGGTCAACCTGCTGACCAAGTTCCTGTTCTACTACCTGAAGAAGAACGAAA
ACATCATCCAGAAGAAGTCCATCGTCGGCTCCTCCCGCCCGTCCGTCCGCGAATACAC
CCTGTCCGAAATCGCCATCAAGATCCCGTCCCTGGAAATCCAGTCCGCCATCCTGGGC
ATCAACGAACACTTCCACTACGTCTAAGGTACC

>MpuUVI
CCATGGAAATCTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGGGCAAGTCCAAGTA
CAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTACTCCTCCAAGACC
AAGGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAACGGCGAATACATCC
TGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCAACGAAAAGTTCTC
CACCACCTCCAACTGCTTCATCCTGAAGGTCGACGAAAACATCGCCAAGACCAAGTTC
CTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATGGCCATCGGCTCCG
CCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCAACCTGCCGAACCT
GAAGACCCAGTCCGCCATCATCAAGATCATCGAACCGAAGGAAGACCTGTTCTTCCGC
CACAAGAACCTGGTCCGCATCGACTCCGAAGAAAACACCAAGAAGGACCTGTCCATCC
TGATCAAGATCATCGAACCGCTGGAAAAGCAGATCAACGCCTTCGACGAACTGATCCT
GTCCGAACAGAAGTCCCTGCAGCACTACCTGAACTACTTCTTCGGCAAGTTCTACCAG
ATCGAACCGTCCCTGTTCCACGACTACAAGCTGGAAAAGATCGCCAAGATCCGCCGCG
GCAAGATCATCAACTCCTTCGACCTGAAGGAAAACCCGGGCGACTACCCGGTCATCTC
CTCCAACACCAAGAACAACGGCATCTTCGGCTACCTGAACTCCTACATGTACGACGGC
GAATACATCACCATCTCCGCCGACGGCGCCTACGCCGGCACCGTCTTCCTGAACAACG
GCAAGTTCTCCATCACCAACGTCTGCTTCATCCTGCTGCTGAACGACAAGGTCAACCT
GCTGACCAAGTTCCTGTTCTACTACCTGAAGAAGAACGAAAACATCATCCAGAAGAAG
TCCATCGTCGGCTCCTCCCGCCCGTCCGTCCGCGAATACACCCTGTCCGAAATCGCCA
TCAAGATCCCGTCCCTGGAAATCCAGTCCGCCATCCTGGGCATCATCGAACCGCTGCA
CAAGAAGATCAACCTGCTGAAGCAGAAGAAGAAGCTGCTGGAAAAGCGCTCCATCTAC
TGCCAGAACCACCTGATCAAGGAAAAGATCAAGGACGAATAGGGTACC

>MpuUVIII
CCATGGAAATCTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGGGCAAGTCCAAGTA
CAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTACTCCTCCAAGACC
AAGGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAACGGCGAATACATCC
TGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCAACGAAAAGTTCTC
CACCACCTCCAACTGCTTCATCCTGAAGGTCGACGAAAACATCGCCAAGACCAAGTTC
CTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATGGCCATCGGCTCCG
CCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCAACCTGCCGAACCT
GAAGACCCAGTCCGCCATCATCAAGATCATCGAACCGCTGGAAAAGCAGATCAACGCC
TTCGACGAACTGATCCTGTCCGAACAGAAGTCCCTGCAGCACTACCTGAACTACTTCC
TGAACAAGCTGGCCTCCATCAACCCGTCCATCTTCAAGAACTACAAGCTGGGCCAGAT
CCTGAACCTGGAAAAGGGCAAGTCCAAGTACAACGCCAAGTACGTCTCCCAGAACATC
GGCATCTACAACCTGTACTCCTCCAAGACCCGCGACCAGGGCATCTTCGGCAAGATCA
ACTCCTACGACTTCAACGGCGAATACATCCTGATCACCACCCACGGCGCCTACGCCGG
CACCGTCAAGTACGTCAACGAAAAGTTCTCCACCACCTCCAACTGCTTCATCCTGAAG
GTCAACGAAAACATCGTCAAGACCAAGTTCCTGTCCTACCTGCTGCTGCTGCAGGAAA
AGACCTTCAACGACATGGCCATCGGCTCCGCCTACGGCTACCTGAAGAACTACAACAT

CAACGACTTCGAAGTCAACCTGCCGAACCTGAAGATCCAGTCCGCCATCCTGGGCATC
ATCGAACCGCTGCACAAGAAGATCAACCTGCTGAAGCAGAAGAAGAAGCTGCTGGAAA
AGCGCTCCATCTACTGCCAGAACCACCTGATCAAGGAAAAGATCAAGGACGAATAGGG
TACC

>MpuT9
CCATGGAAATCTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGGGCAAGTCCAAGTA
CAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTACTCCTCCAAGACC
AAGGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAACGGCGAATACATCC
TGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCAACGAAAAGTTCTC
CACCACCTCCAACTGCTTCATCCTGAAGGTCAACGAAAACATCGTCAAGACCAAGTTC
CTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATGGCCATCGGCTCCG
CCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCAACCTGCCGAACCT
GAAGATCCAGTCCGCCATCATCAAGATCATCGAACCGAAGGAAGACCTGTTCTTCCGC
CACAAGAACCTGGTCCGCATCGACTCCGAAGAAAACACCAAGAAGGACCTGTCCATCC
TGATCAAGATCATCGAACCGCTGGAAAAGCAGATCAACGCCTTCGACGAACTGATCCT
GTCCGAACAGAAGTCCCTGCAGCACTACCTGAACTACTTCCTGAACAAGCTGGCCTCC
ATCAACCCGTCCATCTTCAAGAACTACAAGCTGGGCGAAATCGCCAAGATCCTGTCCG
GCAAGACCCCGTCCACCGCCAAGAAGGAACTGTGGAAGAAGGAAATCCCGTTCTTCGG
CCCGGGCGACCTGGACAACATGGTCCCGAAGCGCTTCATCACCTTCAACGAAAAGATG
ATCAAGCGCTCCGGCACCATCCTGTTCTCCTCCGCCGCCACCATCGGCAAGGTCGGCA
TCCTGGACAACCTGTCCTGGTTCAACCAGCAGATCACCTCCATCGAAGCCAACAACAA
CTACGTCATGGACAAGTTCCTGTTCTTCCTGCTGAAGAAGATCTCCTCCAAGATCAAG
TTCGAAAACTCCTCCGGCACCATCTTCCCGACCATCAAGAAGAAGTACTTCGAAAACT
TCACCCTGGAAATCCCGAACCTGAAGACCCAGTCCGCCATCCTGGGCATCAACGAACA
CTTCCACTACGTCTAAGGTACC

>MpuT10
CCATGGAAATCTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGGGCAAGTCCAAGTA
CAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTACTCCTCCAAGACC
AAGGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAACGGCGAATACATCC
TGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCAACGAAAAGTTCTC
CACCACCTCCAACTGCTTCATCCTGAAGGTCGACGAAAACATCGCCAAGACCAAGTTC
CTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATGGCCATCGGCTCCG
CCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCAACCTGCCGAACCT
GAAGACCCAGTCCGCCATCATCAAGATCATCGAACCGCTGGAAAAGCAGATCAACGCC
TTCGACGAACTGATCCTGTCCGAACAGAAGTCCCTGCAGCACTACCTGAACTACTTCC
TGAACAAGCTGGCCTCCATCAACCCGTCCATCTTCAAGAACTACAAGCTGGGCCAGAT
CCTGAACCTGGAAAAGGGCAAGTCCAAGTACAACGCCAAGTACGTCTCCCAGAACATC
GGCATCTACAACCTGTACTCCTCCAAGACCCGCGACCAGGGCATCTTCGGCAAGATCA
ACTCCTACGACTTCAACGGCGAATACATCCTGATCACCACCCACGGCGCCTACGCCGG
CACCGTCAAGTACGTCAACGAAAAGTTCTCCACCACCTCCAACTGCTTCATCCTGAAG
GTCAACGAAAACATCGTCAAGACCAAGTTCCTGTCCTACCTGCTGCTGCTGCAGGAAA
AGACCTTCAACGACATGGCCATCGGCTCCGCCTACGGCTACCTGAAGAACTACAACAT
CAACGACTTCGAAGTCAACCTGCCGAACCTGAAGATCCAGTCCGCCATCCTGGGCATC
AACGAACACTTCCACTACGTCTAAGGTACC

>MpuT11
CCATGGAAATCTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGGGCAAGTCCAAGTA
CAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTACTCCTCCAAGACC
AAGGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAACGGCGAATACATCC
TGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCAACGAAAAGTTCTC
CACCACCTCCAACTGCTTCATCCTGAAGGTCAACGAAAACATCGTCAAGACCAAGTTC
CTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATGGCCATCGGCTCCG
CCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCAACCTGCCGAACCT
GAAGATCCAGTCCGCCATCATCAAGATCATCGAACCGCTGGAAAAGCAGATCAACGCC
TTCGACGAACTGATCCTGTCCGAACAGAAGTCCCTGCAGCACTACCTGAACTACTTCC
TGAACAAGCTGGCCTCCATCAACCCGTCCATCTTCAAGAACTACAAGCTGGGCGAAAT
CGCCAAGATCCTGTCCGGCAAGACCCCGTCCACCGCCAAGAAGGAACTGTGGAAGAAG
GAAATCCCGTTCTTCGGCCCGGGCGACCTGGACAACATGGTCCCGAAGCGCTTCATCA

CCTTCAACGAAAAGATGATCAAGCGCTCCGGCACCATCCTGTTCTCCTCCGCCGCCAC
CATCGGCAAGGTCGGCATCCTGGACAACCTGTCCTGGTTCAACCAGCAGATCACCTCC
ATCGAAGCCAACAACAACTACGTCATGGACAAGTTCCTGTTCTTCCTGCTGAAGAAGA
TCTCCTCCAAGATCAAGTTCGAAAACTCCTCCGGCACCATCTTCCCGACCATCAAGAA
GAAGTACTTCGAAAACTTCACCCTGGAAATCCCGAACCTGAAGACCCAGTCCGCCATC
CTGGGCATCATCGAACCGCTGCACAAGAAGATCAACCTGCTGAAGCAGAAGAAGAAGC
TGCTGGAAAAGCGCTTCATCTACTACCAGAACCACCTGATCAAGGAAAAGATCAAGGA
CGAATAGGGTACC

>MpuT19
CCATGGAAATCTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGGGCAAGTCCAAGTA
CAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTACTCCTCCAAGACC
AAGGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAACGGCGAATACATCC
TGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCAACGAAAAGTTCTC
CACCACCTCCAACTGCTTCATCCTGAAGGTCAACGAAAACATCGTCAAGACCAAGTTC
CTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATGGCCATCGGCTCCG
CCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCAACCTGCCGAACCT
GAAGATCCAGTCCGCCATCATCAAGATCATCGAACCGCTGGAAAAGCAGATCAACGCC
TTCGACGAACTGATCCTGTCCGAACAGAAGTCCCTGCAGCACTACCTGAACTACTTCC
TGAACAAGCTGGCCTCCATCAACCCGTCCATCTTCAAGAACTACAAGCTGGGCGAAAT
CGCCAAGATCCTGTCCGGCAAGACCCCGTCCACCGCCAAGAAGGAACTGTGGAAGAAG
GAAATCCCGTTCTTCGGCCCGGGCGACCTGGACAACATGGTCCCGAAGCGCTTCATCA
CCTTCAACGAAAAGATGATCAAGCGCTCCGGCACCATCCTGTTCTCCTCCGCCGCCAC
CATCGGCAAGGTCGGCATCCTGGACAACCTGTCCTGGTTCAACCAGCAGATCACCTCC
ATCGAAGCCAACAACAACTACGTCATGGACAAGTTCCTGTTCTTCCTGCTGAAGAAGA
TCTCCTCCAAGATCAAGTTCGAAAACTCCTCCGGCACCATCTTCCCGACCATCAAGAA
GAAGTACTTCGAAAACTTCACCCTGGAAATCCCGAACCTGAAGACCCAGTCCGCCATC
CTGGGCATCAACGAACACTTCCACTACGTCTAAGGTACC

>MpuT22
CCATGGAAATCTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGGGCAAGTCCAAGTA
CAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTACTCCTCCAAGACC
AAGGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAACGGCGAATACATCC
TGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCAACGAAAAGTTCTC
CACCACCTCCAACTGCTTCATCCTGAAGGTCGACGAAAACATCGCCAAGACCAAGTTC
CTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATGGCCATCGGCTCCG
CCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCAACCTGCCGAACCT
GAAGACCCAGTCCGCCATCATCAAGATCATCGAACCGAAGGAAGACCTGTTCTTCCGC
CACAAGAACCTGGTCCGCATCGACTCCGAAGAAAACACCAAGAAGGACCTGTCCATCC
TGATCAAGATCATCGAACCGCTGGAAAAGCAGATCAACGCCTTCGACGAACTGATCCT
GTCCGAACAGAAGTCCCTGCAGCACTACCTGAACTACTTCCTGAACAAGCTGGCCTCC
ATCAACCCGTCCATCTTCAAGAACTACAAGCTGGGCCAGATCCTGAACCTGGAAAAGG
GCAAGTCCAAGTACAACGCCAAGTACGTCTCCCAGAACATCGGCATCTACAACCTGTA
CTCCTCCAAGACCCGCGACCAGGGCATCTTCGGCAAGATCAACTCCTACGACTTCAAC
GGCGAATACATCCTGATCACCACCCACGGCGCCTACGCCGGCACCGTCAAGTACGTCA
ACGAAAAGTTCTCCACCACCTCCAACTGCTTCATCCTGAAGGTCAACGAAAACATCGT
CAAGACCAAGTTCCTGTCCTACCTGCTGCTGCTGCAGGAAAAGACCTTCAACGACATG
GCCATCGGCTCCGCCTACGGCTACCTGAAGAACTACAACATCAACGACTTCGAAGTCA
ACCTGCCGAACCTGAAGATCCAGTCCGCCATCCTGGGCATCAACGAACACTTCCACTA
CGTCTAAGGTACC

Bibliography

Abudayyeh, O. O. *et al.* (2016) 'C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector', *Science (New York, N.Y.)*, 353(6299), p. aaf5573. doi: 10.1126/science.aaf5573 [doi].

Acevedo-Whitehouse, K. *et al.* (2003) 'Inbreeding: Disease susceptibility in California sea lions', *Nature*, 422(6927), p. 35. doi: 10.1038/422035a [doi].

Alder, M. N. *et al.* (2005) 'Diversity and function of adaptive immune receptors in a jawless vertebrate', *Science (New York, N.Y.)*, 310(5756), pp. 1970–1973. doi: 310/5756/1970 [pii].

Allen, C. D. *et al.* (2007) 'Imaging of germinal center selection events during affinity maturation', *Science (New York, N.Y.)*, 315(5811), pp. 528–531. doi: 1136736 [pii].

Altermatt, F. and Ebert, D. (2008) 'Genetic diversity of Daphnia magna populations enhances resistance to parasites', *Ecology Letters*, 11(9), pp. 918–928. doi: 10.1111/j.1461-0248.2008.01203.x [doi].

Altschul, S. F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403–410. doi: 10.1016/S0022-2836(05)80360-2.

Antia, R. *et al.* (2003) 'The role of evolution in the emergence of infectious diseases', *Nature*, 426(6967), pp. 658–661. doi: 10.1038/nature02104 [doi].

Arber, W. and Linn, S. (1969) 'DNA Modification and Restriction', *Annual Review of Biochemistry*, 38, pp. 467–500. Available at: http://www.annualreviews.org/doi/pdf/10.1146/annurev.bi.38.070169.002343 (Accessed: 15 September 2017).

Artimo, P. *et al.* (2012) 'ExPASy: SIB bioinformatics resource portal', *Nucleic Acids Research*, 40(W1). doi: 10.1093/nar/gks400.

Ashby, B. and King, K. C. (2015) 'Diversity and the maintenance of sex by parasites', *Journal of Evolutionary Biology*, 28(3), pp. 511–520. doi: 10.1111/jeb.12590 [doi].

Atanasiu, C. *et al.* (2002) 'Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme', *Nucleic acids research*, 30(18), pp. 3936–3944.

Baer, B. and Schmid-Hempel, P. (1999) 'Experimental variation in polyandry affects parasite loads and fitness in a bumble-bee', *Nature*, 397(6715), pp.

151–154.

Baer, B. and Schmid-Hempel, P. (2001) ‘Unexpected consequences of polyandry for parasitism and fitness in the bumblebee, Bombus terrestris’, *Evolution; international journal of organic evolution*, 55(8), pp. 1639–1643.

Baer, B. and Schmid-Hempel, P. (2003) ‘Bumblebee workers from different sire groups vary in suscetibility to parasite infection’, *Ecology Letters*, 6(2), pp. 106–110.

Barrangou, R. *et al.* (2007) ‘CRISPR provides acquired resistance against viruses in prokaryotes’, *Science (New York, N.Y.)*, 315(5819), pp. 1709–1712. doi: 315/5819/1709 [pii].

Belkaid, Y. *et al.* (2002) ‘CD4+CD25+ regulatory T cells control Leishmania major persistence and immunity’, *Nature*, 420(6915), pp. 502–507. doi: 10.1038/nature01152 [doi].

Benchling (2015) *Benchling · Better tools, faster research*. Available at: https://benchling.com/ (Accessed: 11 September 2017).

Bikard, D. and Marraffini, L. A. (2012) ‘Innate and adaptive immunity in bacteria: mechanisms of programmed genetic variation to fight bacteriophages’, *Current opinion in immunology*, 24(1), pp. 15–20. doi: 10.1016/j.coi.2011.10.005 [doi].

Biswas, A. *et al.* (2013) ‘CRISPRTarget: bioinformatic prediction and analysis of crRNA targets.’, *RNA biology*. Taylor & Francis, 10(5), pp. 817–27. doi: 10.4161/rna.24046.

Boehm, T. *et al.* (2012) ‘VLR-based adaptive immunity’, *Annual Review of Immunology*, 30, pp. 203–220. doi: 10.1146/annurev-immunol-020711-075038 [doi].

Bondy-Denomy, J. *et al.* (2013) ‘Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system’, *Nature*, 493(7432), pp. 429–432. doi: 10.1038/nature11723 [doi].

Bondy-Denomy, J. *et al.* (2015) ‘Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins’, *Nature*, 526(7571), pp. 136–139. doi: 10.1038/nature15254 [doi].

Boulinier, T. and Staszewski, V. (2008) ‘Maternal transfer of antibodies: raising immuno-ecology issues’, *Trends in ecology & evolution*, 23(5), pp. 282–288. doi: 10.1016/j.tree.2007.12.006 [doi].

Breitbart, M. and Rohwer, F. (2005) ‘Here a virus, there a virus, everywhere the

same virus?', *Trends in Microbiology*, 13(6), pp. 278–284. doi: 10.1016/j.tim.2005.04.003.

Brocchi, M., de Vasconcelos, A. T. R. and Zaha, A. (2007) 'Restriction-modification systems in Mycoplasma spp', *Genetics and Molecular Biology*, 30(SUPPL. 1), pp. 236–244. doi: 10.1590/S1415-47572007000200011.

Brouns, S. J. *et al.* (2008) 'Small CRISPR RNAs guide antiviral defense in prokaryotes', *Science (New York, N.Y.)*, 321(5891), pp. 960–964. doi: 10.1126/science.1159689 [doi].

Buckling, A. and Rainey, P. B. (2002) 'Antagonistic coevolution between a bacterium and a bacteriophage', *Proceedings.Biological sciences*, 269(1494), pp. 931–936. doi: 10.1098/rspb.2001.1945 [doi].

Burckhardt, J. *et al.* (1981) 'Complexes formed between the restriction endonuclease EcoK and heteroduplex DNA', *Journal of Molecular Biology*, 153(2), pp. 425–440. doi: 10.1016/0022-2836(81)90287-4.

Cady, K. C. *et al.* (2012) 'The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages', *Journal of Bacteriology*, 194(21), pp. 5728–5738. doi: 10.1128/JB.01184-12.

Cady, K. C. and O'Toole, G. A. (2011) 'Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins', *Journal of Bacteriology*, 193(14), pp. 3433–3445. doi: 10.1128/JB.01411-10.

Ceyssens, P. J. *et al.* (2009) 'Comparative analysis of the widespread and conserved PB1-like viruses infecting Pseudomonas aeruginosa', *Environmental Microbiology*, 11(11), pp. 2874–2883. doi: 10.1111/j.1462-2920.2009.02030.x.

Chabas, H. *et al.* (2016) 'Immigration of susceptible hosts triggers the evolution of alternative parasite defence strategies', *Proceedings.Biological sciences*, 283(1837), p. 10.1098/rspb.2016.0721. doi: 10.1098/rspb.2016.0721 [doi].

Chibeu, A. *et al.* (2009) 'The adsorption of Pseudomonas aeruginosa bacteriophage φKMV is dependent on expression regulation of type IV pili genes', *FEMS Microbiology Letters*, 296(2), pp. 210–218. doi: 10.1111/j.1574-6968.2009.01640.x.

Chuang, J. H. and Li, H. (2004) 'Functional bias and spatial organization of genes in mutational hot and cold regions in the human genome', *PLoS biology*, 2(2), p. E29. doi: 10.1371/journal.pbio.0020029 [doi].

Clark, K. *et al.* (2016) 'GenBank.', *Nucleic acids research*. Oxford University Press, 44(D1), pp. D67-72. doi: 10.1093/nar/gkv1276.

Cowan, G. M., Gann, A. A. F. and Murray, N. E. (1989) 'Conservation of complex DNA recognition domains between families of restriction enzymes', *Cell*, 56(1), pp. 103–109. doi: 10.1016/0092-8674(89)90988-4.

Datta, A. and Jinks-Robertson, S. (1995) 'Association of increased spontaneous mutation rates with high levels of transcription in yeast', *Science (New York, N.Y.)*, 268(5217), pp. 1616–1619.

Decaestecker, E. *et al.* (2007) 'Host-parasite "Red Queen" dynamics archived in pond sediment', *Nature*, 450(7171), pp. 870–873. doi: nature06291 [pii].

Deveau, H. *et al.* (2008) 'Phage response to CRISPR-encoded resistance in Streptococcus thermophilus', *Journal of Bacteriology*, 190(4), pp. 1390–1400. doi: JB.01412-07 [pii].

Ding, S. W. and Voinnet, O. (2007) 'Antiviral immunity directed by small RNAs', *Cell*, 130(3), pp. 413–426. doi: S0092-8674(07)00977-4 [pii].

Dong, J. *et al.* (2015) 'Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching', *Nature*, 525(7567), pp. 134–139. doi: 10.1038/nature14970 [doi].

Doulatov, S. *et al.* (2004) 'Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements', *Nature*, 431(7007), pp. 476–481. doi: 10.1038/nature02833 [doi].

Dryden, D. T. F. *et al.* (1997) 'The in vitro assembly of the EcoKI type I DNA restriction/modification enzyme and its in vivo implications', *Biochemistry*. King & Murray, 36(5), pp. 1065–1076. doi: 10.1021/bi9619435.

Dupuis, M.-È. *et al.* (2013) 'CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance.', *Nature communications*, 4(May), p. 2087. doi: 10.1038/ncomms3087.

Dybvig, K., Sitaraman, R. and French, C. T. (1998) 'A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements', *Proceedings of the National Academy of Sciences*, 95(November), pp. 13923–13928. doi: 10.1073/pnas.95.23.13923.

Dybvig, K. and Yu, H. (1994) 'Regulation of a restriction and modification system via DNA inversion in Mycoplasma pulmonis', *Molecular Microbiology*, 12(4), pp. 547–560. doi: 10.1111/j.1365-2958.1994.tb01041.x.

Early, P. *et al.* (1980) ‘An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH’, *Cell*, 19(4), pp. 981–992. doi: 0092-8674(80)90089-6 [pii].

East-Seletsky, A. *et al.* (2016) ‘Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection’, *Nature*, 538(7624), pp. 270–273. doi: 10.1038/nature19802 [doi].

Elton, C. S. (1958) *The Ecology of Invasions by Animals and Plants*. London: Methuen.

Ershova, A. S. *et al.* (2015) ‘Role of restriction-modification systems in prokaryotic evolution and ecology’, *Biochemistry (Moscow)*, 80(10), pp. 1373–1386. doi: 10.1134/S0006297915100193.

Finch, C. E. and Crimmins, E. M. (2004) ‘Inflammatory exposure and historical changes in human life-spans’, *Science (New York, N.Y.)*, 305(5691), pp. 1736–1739. doi: 10.1126/science.1092556 [doi].

Forsberg, L. A., Gisselsson, D. and Dumanski, J. P. (2017) ‘Mosaicism in health and disease - clones picking up speed’, *Nature reviews.Genetics*, 18(2), pp. 128–142. doi: 10.1038/nrg.2016.145 [doi].

Foster, P. L. (2007) ‘Stress-induced mutagenesis in bacteria’, *Critical reviews in biochemistry and molecular biology*, 42(5), pp. 373–397. doi: 782797248 [pii].

Gann, A. A. F. *et al.* (1987) ‘Reassortment of DNA recognition domains and the evolution of new specificities’, *Molecular Microbiology*, 1(3), pp. 13–22. doi: 10.1111/j.1365-2958.1987.tb00521.x.

Garneau, J. E. *et al.* (2010) ‘The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA’, *Nature*, 468(7320), pp. 67–71. doi: 10.1038/nature09523 [doi].

Gemmill, A. W., Viney, M. E. and Read, A. F. (1997) ‘Host immune status determines sexuality in a parasitic nematode’, *Evolution*, 51(2), pp. 393–401. doi: 10.2307/2411111.

Gerber, M. *et al.* (2014) ‘Selective packaging of the influenza A genome and consequences for genetic reassortment’, *Trends in microbiology*, 22(8), pp. 446–455. doi: 10.1016/j.tim.2014.04.001.

Giraud, A. *et al.* (2001) ‘Costs and benefits of high mutation rates: Adaptive evolution of bacteria in the mouse gut’, *Science*, 291(5513), pp. 2606–2608. doi: 10.1126/science.1056421.

Gitlin, A. D. *et al.* (2015) 'HUMORAL IMMUNITY. T cell help controls the speed of the cell cycle in germinal center B cells', *Science (New York, N.Y.)*, 349(6248), pp. 643–646. doi: 10.1126/science.aac4919 [doi].

Gitlin, A. D., Shulman, Z. and Nussenzweig, M. C. (2014) 'Clonal selection in the germinal centre by regulated proliferation and hypermutation', *Nature*, 509(7502), pp. 637–640. doi: 10.1038/nature13300 [doi].

Goldberg, G. W. *et al.* (2014) 'Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting', *Nature*, 514(7524), pp. 633–637. doi: 10.1038/nature13637 [doi].

Gomez, P. and Buckling, A. (2011) 'Bacteria-phage antagonistic coevolution in soil', *Science (New York, N.Y.)*, 332(6025), pp. 106–109. doi: 10.1126/science.1198767 [doi].

Graham, A. L. *et al.* (2010) 'Fitness correlates of heritable variation in antibody responsiveness in a wild mammal', *Science (New York, N.Y.)*, 330(6004), pp. 662–665. doi: 10.1126/science.1194878 [doi].

Graham, A. L., Allen, J. E. and Read, A. F. (2005) 'Evolutionary causes and consequences of immunopathology', *Annual Review of Ecology Evolution and Systematics*, 36, pp. 373–397. doi: 10.1146/annurev.ecolsys.36.102003.152622.

Griffith, F. (1928) 'The Significance of Pneumococcal Types', *The Journal of hygiene*, 27(2), pp. 113–159.

Grissa, I., Vergnaud, G. and Pourcel, C. (2007a) 'CRISPRFinder: a web tool to identify clustered regularly interspace short palindromic repeats', *Nucleic Acids Research*, 35(Web Server issue), pp. 52–57. doi: 10.1093/nar/gkn228.

Grissa, I., Vergnaud, G. and Pourcel, C. (2007b) 'The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats', *BMC Bioinformatics*, 8(1), p. 172. doi: 10.1186/1471-2105-8-172.

Gumulak-Smith, J. *et al.* (2001) 'Variations in the surface proteins and restriction enzyme systems of Mycoplasma pulmonis in the respiratory tract of infected rats', *Molecular Microbiology*, 40(4), pp. 1037–1044. doi: 10.1046/j.1365-2958.2001.02464.x.

Hamilton, A. J. and Baulcombe, D. C. (1999) 'A species of small antisense RNA in posttranscriptional gene silencing in plants', *Science (New York, N.Y.)*, 286(5441), pp. 950–952. doi: 7953 [pii].

Hamilton, W. D., Axelrod, R. and Tanese, R. (1990) 'Sexual reproduction as an adaptation to resist parasites (a review)', *Proceedings of the National Academy of Sciences of the United States of America*, 87(9), pp. 3566–3573.

Hamilton, W. D. and Zuk, M. (1982) 'Heritable true fitness and bright birds: a role for parasites?', *Science (New York, N.Y.)*, 218(4570), pp. 384–387.

Hardison, R. C. *et al.* (2003) 'Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution', *Genome research*, 13(1), pp. 13–26. doi: 10.1101/gr.844103 [doi].

Herman, R. K. and Dworkin, N. B. (1971) 'Effect of gene induction on the rate of mutagenesis by ICR-191 in Escherichia coli', *Journal of Bacteriology*, 106(2), pp. 543–550.

Heussler, G. E. *et al.* (2015) 'Clustered regularly interspaced short palindromic repeat-dependent, biofilm-specific death of pseudomonas aeruginosa mediated by increased expression of phage-related genes', *mBio*, 6(3), pp. 1–13. doi: 10.1128/mBio.00129-15.

Holz, C. L. *et al.* (2012) 'RNA interference against animal viruses: how morbilliviruses generate extended diversity to escape small interfering RNA control', *Journal of virology*, 86(2), pp. 786–795. doi: 10.1128/JVI.06210-11 [doi].

Horns, F. *et al.* (2016) 'Lineage tracing of human B cells reveals the in vivo landscape of human antibody class switching', *eLife*, 5, p. 10.7554/eLife.16578. doi: 10.7554/eLife.16578 [doi].

van Houte, S. *et al.* (2016) 'The diversity-generating benefits of a prokaryotic adaptive immune system', *Nature*. Nature Publishing Group, 532(7599), pp. 385–388. doi: 10.1038/nature17436.

van Houte, S., Buckling, A. and Westra, E. R. (2016) 'Evolutionary Ecology of Prokaryotic Immune Mechanisms', *Microbiology and Molecular Biology Reviews*, 80(3), pp. 745–763.

Hyatt, D. *et al.* (2010) 'Prodigal: prokaryotic gene recognition and translation initiation site identification', *BMC Bioinformatics*, 11(1), p. 119. doi: 10.1186/1471-2105-11-119.

Hynes, A. P., Villion, M. and Moineau, S. (2014) 'Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages', *Nature Communications*. Nature Publishing Group, 5(May), pp. 1–6. doi:

10.1038/ncomms5399.

Insilico (2017) *Ligation calculator*. Available at: http://www.insilico.uni-duesseldorf.de/Lig_Input.html (Accessed: 11 September 2017).

Ishikawa, K., Fukuda, E. and Kobayashi, I. (2010) 'Conflicts Targeting Epigenetic Systems and Their Resolution by Cell Death: Novel Concepts for Methyl-Specific and Other Restriction Systems', *DNA Research*, 17, pp. 325–342. doi: 10.1093/dnares/dsq027.

Janscak, P. *et al.* (1999) 'DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes', *EMBO Journal*, 18(9), pp. 2638–2647. doi: 10.1093/emboj/18.9.2638.

Jiang, F. *et al.* (2016) 'Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage', *Science (New York, N.Y.)*, 351(6275), pp. 867–871. doi: 10.1126/science.aad8282 [doi].

Jiang, N. *et al.* (2011) 'Determinism and stochasticity during maturation of the zebrafish antibody repertoire', *Proceedings of the National Academy of Sciences of the United States of America*, 108(13), pp. 5348–5353. doi: 10.1073/pnas.1014277108 [doi].

Kamiya, T. *et al.* (2014) 'A quantitative review of MHC-based mating preference: the role of diversity and dissimilarity', *Molecular ecology*, 23(21), pp. 5151–5163. doi: 10.1111/mec.12934 [doi].

Kan, N. C. *et al.* (1979) 'The nucleotide sequence recognized by the Escherichia coli K12 restriction and modification enzymes', *Journal of Molecular Biology*, 130(2), pp. 191–209. doi: 10.1016/0022-2836(79)90426-1.

Keesing, F. *et al.* (2010) 'Impacts of biodiversity on the emergence and transmission of infectious diseases', *Nature*, 468(7324), pp. 647–652. doi: 10.1038/nature09575 [doi].

Keesing, F., Holt, R. D. and Ostfeld, R. S. (2006) 'Effects of species diversity on disease risk', *Ecology Letters*, 9(4), pp. 485–498. doi: ELE885 [pii].

Kelly, L. A. *et al.* (2015) 'The Phyre2 web portal for protein modelling, prediction, and analysis', *Nature Protocols*, 10(6), pp. 845–858. doi: 10.1038/nprot.2015-053.

Ketting, R. F. (2011) 'The many faces of RNAi', *Developmental cell*, 20(2), pp. 148–161. doi: 10.1016/j.devcel.2011.01.012 [doi].

Kim, J.-S. *et al.* (2005) 'Crystal structure of DNA sequence specificity subunit of

a type I restriction-modification enzyme and its functional implications.', *Proceedings of the National Academy of Sciences of the United States of America*, 102(9), pp. 3248–53. doi: 10.1073/pnas.0409851102.

Kim, M. S. *et al.* (2015) 'Crystal structure of the V(D)J recombinase RAG1-RAG2', *Nature*, 518(7540), pp. 507–511. doi: 10.1038/nature14174 [doi].

King, K. C. *et al.* (2009) 'The geographic mosaic of sex and the Red Queen', *Current biology : CB*, 19(17), pp. 1438–1441. doi: 10.1016/j.cub.2009.06.062 [doi].

King, K. C. and Lively, C. M. (2009) 'Geographic variation in sterilizing parasite species and the Red Queen', *Oikos*, 118(9), pp. 1416–1420. doi: 10.1111/j.1600-0706.2009.17476.x.

Kobasa, D. *et al.* (2007) 'Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus', *Nature*, 445(7125), pp. 319–323. doi: nature05495 [pii].

Kobayashi, I. (2001) 'Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution', *Nucleic Acids Research*, 29(18), pp. 3742–3756. doi: 10.1093/nar/29.18.3742.

Kojima, K. K. *et al.* (2016) 'Population Evolution of Helicobacter pylori through Diversification in DNA Methylation and Interstrain Sequence Homogenization', *Molecular Biology and Evolution*, 33(11), pp. 2848–2859. doi: 10.1093/molbev/msw162.

Korona, R. and Levin, B. R. (1993) 'Phage-Mediated Selection and the Evolution and Maintenance of Restriction-Modification', *Evolution*, 47(2), pp. 556–575. doi: 10.2307/2410071.

Krüger, D. H. and Bickle, T. a (1983) 'Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts.', *Microbiological reviews*, 47(3), pp. 345–360. Available at: http://mmbr.asm.org/content/47/3/345.full.pdf (Accessed: 7 September 2017).

Kulik, E. M. and Bickle, T. A. (1996) 'Regulation of the activity of the type IC EcoR124I restriction enzyme.', *Journal of molecular biology*, 264(5), pp. 891–906. doi: 10.1006/jmbi.1996.0685.

Kuraoka, M. *et al.* (2016) 'Complex Antigens Drive Permissive Clonal Selection in Germinal Centers', *Immunity*, 44(3), pp. 542–552. doi: 10.1016/j.immuni.2016.02.010 [doi].

Labrie, S. J., Samson, J. E. and Moineau, S. (2010) 'Bacteriophage resistance mechanisms', *Nature Reviews Microbiology*, 8(5), pp. 317–327. doi: 10.1038/nrmicro2315.

Lafforgue, G. *et al.* (2011) 'Tempo and mode of plant RNA virus escape from RNA interference-mediated resistance', *Journal of virology*, 85(19), pp. 9686–9695. doi: 10.1128/JVI.05326-11 [doi].

LeClerc, J. E. *et al.* (1996) 'High mutation frequencies among Escherichia coli and Salmonella pathogens', *Science (New York, N.Y.)*, 274(5290), pp. 1208–1211.

Lee, H. *et al.* (2012) 'Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing', *Proceedings of the National Academy of Sciences of the United States of America*, 109(41), pp. E2774-83. doi: 10.1073/pnas.1210309109 [doi].

Leinders-Zufall, T. *et al.* (2004) 'MHC class I peptides as chemosensory signals in the vomeronasal organ', *Science (New York, N.Y.)*, 306(5698), pp. 1033–1037. doi: 306/5698/1033 [pii].

Levin, B. R. *et al.* (2013) 'The population and evolutionary dynamics of phage and bacteria with CRISPR-mediated immunity', *PLoS genetics*, 9(3), p. e1003312. doi: 10.1371/journal.pgen.1003312 [doi].

Levin, B. R., Antonovics, J. and Sharma, H. (1988) 'Frequency-Dependent Selection in Bacterial Populations', *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences*, 319(1196), pp. 459–472.

Li, H., Li, W. X. and Ding, S. W. (2002) 'Induction and suppression of RNA silencing by an animal virus', *Science (New York, N.Y.)*, 296(5571), pp. 1319–1321. doi: 10.1126/science.1070948 [doi].

Li, Y. *et al.* (2013) 'RNA interference functions as an antiviral immunity mechanism in mammals', *Science (New York, N.Y.)*, 342(6155), pp. 231–234. doi: 10.1126/science.1241911 [doi].

Lindberg, R. B. and Latta, R. L. (1974) 'Phage typing of Pseudomonas aeruginosa: clinical and epidemiologic considerations.', *Journal of Infectious Diseases*, 130 Suppl, pp. 33–42. Available at: http://www.jstor.org/stable/30081952 (Accessed: 13 September 2017).

Lively, C. M. (1987) 'Evidence from a New-Zealand Snail for the Maintenance of Sex by Parasitism', *Nature*, 328(6130), pp. 519–521. doi:

10.1038/328519a0.

Lively, C. M. (2010a) ‘A review of Red Queen models for the persistence of obligate sexual reproduction’, *The Journal of heredity*, 101 Suppl, pp. S13-20. doi: 10.1093/jhered/esq010 [doi].

Lively, C. M. (2010b) ‘The effect of host genetic diversity on disease spread’, *The American Naturalist*, 175(6), pp. E149-52. doi: 10.1086/652430 [doi].

Lively, C. M., Craddock, C. and Vrijenhoek, R. C. (1990) ‘Red Queen Hypothesis Supported by Parasitism in Sexual and Clonal Fish’, *Nature*, 344(6269), pp. 864–866. doi: 10.1038/344864a0.

Lively, C. M. and Dybdahl, M. F. (2000) ‘Parasite adaptation to locally common host genotypes’, *Nature*, 405(6787), pp. 679–681. doi: 10.1038/35015069 [doi].

Lively, C. M. and Morran, L. T. (2014) ‘The ecology of sexual reproduction’, *Journal of Evolutionary Biology*, 27(7), pp. 1292–1303. doi: 10.1111/jeb.12354 [doi].

Lo, A. and Qi, L. (2017) ‘Genetic and epigenetic control of gene expression by CRISPR–Cas systems’, *F1000Research*, 6(May), p. 747. doi: 10.12688/f1000research.11113.1.

Lu, R. *et al.* (2005) ‘Animal virus replication and RNAi-mediated antiviral silencing in Caenorhabditis elegans’, *Nature*, 436(7053), pp. 1040–1043. doi: nature03870 [pii].

Luijckx, P. *et al.* (2011) ‘Cloning of the unculturable parasite Pasteuria ramosa and its Daphnia host reveals extreme genotype-genotype interactions’, *Ecology Letters*, 14(2), pp. 125–131. doi: 10.1111/j.1461-0248.2010.01561.x [doi].

Luijckx, P. *et al.* (2013) ‘A matching-allele model explains host resistance to parasites’, *Current biology : CB*, 23(12), pp. 1085–1088. doi: 10.1016/j.cub.2013.04.064 [doi].

Lumley, A. J. *et al.* (2015) ‘Sexual selection protects against extinction’, *Nature*, 522(7557), pp. 470–473. doi: 10.1038/nature14419 [doi].

Lyczak, J. B., Cannon, C. L. and Pier, G. B. (2000) ‘Establishment of Pseudomonas aeruginosa infection: Lessons from a versatile opportunist’, *Microbes and Infection*, 2(9), pp. 1051–1060. doi: 10.1016/S1286-4579(00)01259-4.

Lynch, M. (2010) ‘Rate, molecular spectrum, and consequences of human

mutation', *Proceedings of the National Academy of Sciences of the United States of America*, 107(3), pp. 961–968. doi: 10.1073/pnas.0912629107 [doi].

Lynch, M. *et al.* (2016) 'Genetic drift, selection and the evolution of the mutation rate', *Nature reviews.Genetics*, 17(11), pp. 704–714. doi: 10.1038/nrg.2016.104 [doi].

Maillard, P. V *et al.* (2013) 'Antiviral RNA interference in mammalian cells', *Science (New York, N.Y.)*, 342(6155), pp. 235–238. doi: 10.1126/science.1241930 [doi].

Maizels, N. (1987) 'Diversity achieved by diverse mechanisms: gene conversion in developing B cells of the chicken', *Cell*, 48(3), pp. 359–360. doi: 0092-8674(87)90182-6 [pii].

Makarova, K. S. *et al.* (2015) 'An updated evolutionary classification of CRISPR-Cas systems', *Nature Reviews Microbiology*, 13(11), pp. 722–736. doi: 10.1038/nrmicro3569 [doi].

Makarova, K. S., Wolf, Y. I. and Koonin, E. V (2013) 'Comparative genomics of defense systems in archaea and bacteria', *Nucleic Acids Research*, 41(8), pp. 4360–4377. doi: 10.1093/nar/gkt157.

Makova, K. D. and Hardison, R. C. (2015) 'The effects of chromatin organization on variation in mutation rates in the genome', *Nature reviews.Genetics*, 16(4), pp. 213–223. doi: 10.1038/nrg3890 [doi].

Makovets, S., Doronina, V. A. and Murray, N. E. (1999) 'Regulation of endonuclease activity by proteolysis prevents breakage of unmodified bacterial chromosomes by type I restriction enzymes', *Genetics*, 96(August), pp. 9757–9762. doi: 10.1073/pnas.96.17.9757.

Manso, A. S. *et al.* (2014) 'A random six-phase switch regulates pneumococcal virulence via global epigenetic changes', *Nature Communications*, 5, p. 5055. doi: 10.1038/ncomms6055.

Martinez, F. *et al.* (2012) 'Ultradeep sequencing analysis of population dynamics of virus escape mutants in RNAi-mediated resistant plants', *Molecular biology and evolution*, 29(11), pp. 3297–3307. doi: 10.1093/molbev/mss135 [doi].

Matic, I. *et al.* (1997) 'Highly variable mutation rates in commensal and pathogenic Escherichia coli', *Science (New York, N.Y.)*, 277(5333), pp. 1833–1834.

Maynard Smith, J. (1971) ‘The origin and maintenance of sex’, in Williams, G. C. (ed.). Chicago: Aldine-Atherton (Group Selection), pp. 163–175.

Maynard Smith, J. (1978) *The Evolution of Sex*. Cambridge, UK: Cambridge University Press.

McDonald, M. J., Rice, D. P. and Desai, M. M. (2016) ‘Sex speeds adaptation by altering the dynamics of molecular evolution’, *Nature*, 531(7593), pp. 233–236. doi: 10.1038/nature17143 [doi].

Meister, J. *et al.* (1993) ‘Macroevolution by transposition: drastic modification of DNA recognition by a type I restriction enzyme following Tn5 transposition.’, *The EMBO journal*, 12(12), pp. 4585–91. doi: 10.1002/j.1460-2075.1993.tb06147.x.

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(7648), pp. 101–104. doi: 10.1038/nature21719 [doi].

Molnar, A. *et al.* (2005) ‘Plant virus-derived small interfering RNAs originate predominantly from highly structured single-stranded viral RNAs’, *Journal of virology*, 79(12), pp. 7812–7818. doi: 79/12/7812 [pii].

Morgan, A. D., Bonsall, M. B. and Buckling, A. (2010) ‘Impact of bacterial mutation rate on coevolutionary dynamics between bacteria and phages’, *Evolution; international journal of organic evolution*, 64(10), pp. 2980–2987. doi: 10.1111/j.1558-5646.2010.01037.x [doi].

Morgan, A. D., Gandon, S. and Buckling, A. (2005) ‘The effect of migration on local adaptation in a coevolving host-parasite system’, *Nature*, 437(7056), pp. 253–256. doi: nature03913 [pii].

Morley, D. *et al.* (2017) ‘Host diversity limits the evolution of parasite local adaptation’, *Molecular Ecology*, 26(7), pp. 1756–1763. doi: 10.1111/mec.13917.

Morran, L. T. *et al.* (2011) ‘Running with the Red Queen: host-parasite coevolution selects for biparental sex’, *Science (New York, N.Y.)*, 333(6039), pp. 216–218. doi: 10.1126/science.1206360 [doi].

Morran, L. T., Parmenter, M. D. and Phillips, P. C. (2009) ‘Mutation load and rapid adaptation favour outcrossing over self-fertilization’, *Nature*, 462(7271), pp. 350–352. doi: 10.1038/nature08496 [doi].

Murray, N. E. (2000) ‘Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle).’, *Microbiology and molecular*

biology reviews : MMBR, 64(2), pp. 412–34. doi: 10.1128/MMBR.64.2.412-434.2000.

Nagamalleswari, E. *et al.* (2017) 'Restriction endonuclease triggered bacterial apoptosis as a mechanism for long time survival', *Nucleic Acids Research*. Oxford University Press, 45(14), pp. 8423–8434. doi: 10.1093/nar/gkx576.

Nobusato, A., Uchiyama, I. and Kobayashi, I. (2000) 'Diversity of restriction – modification gene homologues in Helicobacter pylori', *Gene*, 259, pp. 89–98. Available at: www.elsevier.com/locate/gene (Accessed: 11 September 2017).

O'Brien, S. J. *et al.* (1985) 'Genetic basis for species vulnerability in the cheetah', *Science (New York, N.Y.)*, 227(4693), pp. 1428–1434.

O'Sullivan, D. *et al.* (2000) 'Navel type I restriction specificities through domain shuffling of HsdS subunits in Lactococcus lactis', *Molecular Microbiology*, 36(4), pp. 866–875. doi: 10.1046/j.1365-2958.2000.01901.x.

Obbard, D. J. *et al.* (2006) 'Natural selection drives extremely rapid evolution in antiviral RNAi genes', *Current biology : CB*, 16(6), pp. 580–585. doi: S0960-9822(06)01208-5 [pii].

Oliver, A. *et al.* (2000) 'High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection', *Science (New York, N.Y.)*, 288(5469), pp. 1251–1254. doi: 8507 [pii].

Paigen, K. and Petkov, P. (2010) 'Mammalian recombination hot spots: properties, control and evolution', *Nature reviews.Genetics*, 11(3), pp. 221–233. doi: 10.1038/nrg2712 [doi].

Pal, C. *et al.* (2007) 'Coevolution with viruses drives the evolution of bacterial mutation rates', *Nature*, 450(7172), pp. 1079–1081. doi: nature06350 [pii].

Pancer, Z. *et al.* (2004) 'Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey', *Nature*, 430(6996), pp. 174–180. doi: 10.1038/nature02740 [doi].

Pappas, L. *et al.* (2014) 'Rapid development of broadly influenza neutralizing antibodies through redundant mutations', *Nature*, 516(7531), pp. 418–422. doi: 10.1038/nature13764 [doi].

Patrick, S. *et al.* (2010) 'Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of Bacteroides fragilis', *Microbiology*, 156(11), pp. 3255–3269. doi: 10.1099/mic.0.042978-0.

Paul, B. G. *et al.* (2015) 'Targeted diversity generation by intraterrestrial archaea and archaeal viruses', *Nature communications*, 6, p. 6585. doi: 10.1038/ncomms7585 [doi].

Pawluk, A. *et al.* (2014) 'A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa', *mBio*, 5(2), pp. e00896-14. doi: 10.1128/mBio.00896-14 [doi].

Pawluk, A., Staals, R. H., *et al.* (2016) 'Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species', *Nature microbiology*, 1(8), p. 16085. doi: 10.1038/nmicrobiol.2016.85 [doi].

Pawluk, A., Amrani, N., *et al.* (2016) 'Naturally Occurring Off-Switches for CRISPR-Cas9', *Cell*, 167(7), p. 1829–1838.e9. doi: S0092-8674(16)31589-6 [pii].

Pettersen, E. F. *et al.* (2004) 'UCSF Chimera - A visualization system for exploratory research and analysis', *Journal of Computational Chemistry*, 25(13), pp. 1605–1612. doi: 10.1002/jcc.20084.

Pleška, M. *et al.* (2016) 'Bacterial autoimmunity due to a restriction-modification system', *Current Biology*, 26(3), pp. 404–409. doi: 10.1016/j.cub.2015.12.041.

Potts, W. K., Manning, C. J. and Wakeland, E. K. (1991) 'Mating patterns in seminatural populations of mice influenced by MHC genotype', *Nature*, 352(6336), pp. 619–621. doi: 10.1038/352619a0 [doi].

Prakash-Cheng, A. and Ryu, J. (1993) 'Delayed expression of in vivo restriction activity following conjugal transfer of Escherichia coli hsd(K) (restriction-modification) genes', *Journal of Bacteriology*, pp. 4905–4906. Available at: http://jb.asm.org/content/175/15/4905.full.pdf (Accessed: 15 September 2017).

Price, V. J. *et al.* (2016) 'CRISPR-Cas and Restriction-Modification Act Additively against Conjugative Antibiotic Resistance Plasmid Transfer in Enterococcus faecalis', *Molecular Biology and Physiology*, 1(3), pp. 1–13. doi: 10.1128/mSphere.00064-16.Editor.

Puigbò, P. *et al.* (2007) 'OPTIMIZER: A web server for optimizing the codon usage of DNA sequences', *Nucleic Acids Research*, 35(SUPPL.2), pp. 126–131. doi: 10.1093/nar/gkm219.

Pumplin, N. and Voinnet, O. (2013) 'RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence',

Nature reviews.Microbiology, 11(11), pp. 745–760. doi: 10.1038/nrmicro3120 [doi].

Qiu, D. *et al.* (2008) 'PBAD-based shuttle vectors for functional analysis of toxic and highly regulated genes in Pseudomonas and Burkholderia spp. and other bacteria', *Applied and Environmental Microbiology*, 74(23), pp. 7422–7426. doi: 10.1128/AEM.01369-08.

R Core Team (2017) 'R', *R Core Team*. doi: 3-900051-14-3.

Rast, J. P. and Litman, G. W. (1994) 'T-cell receptor gene homologs are present in the most primitive jawed vertebrates', *Proceedings of the National Academy of Sciences of the United States of America*, 91(20), pp. 9248–9252.

Rauch, B. J. *et al.* (2017) 'Inhibition of CRISPR-Cas9 with Bacteriophage Proteins', *Cell*, 168(1–2), p. 150–158.e10. doi: S0092-8674(16)31683-X [pii].

Reusch, T. B. *et al.* (2001) 'Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism', *Nature*, 414(6861), pp. 300–302. doi: 10.1038/35104547 [doi].

Rhoads, A. and Au, K. F. (2015) 'PacBio Sequencing and Its Applications', *Genomics, Proteomics and Bioinformatics*, pp. 278–289. doi: 10.1016/j.gpb.2015.08.002.

Richter, C. *et al.* (2014) 'Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer', *Nucleic acids research*, 42(13), pp. 8516–8526. doi: 10.1093/nar/gku527 [doi].

van Rij, R. P. *et al.* (2006) 'The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster', *Genes & development*, 20(21), pp. 2985–2995. doi: 20/21/2985 [pii].

Roberts, R. J. *et al.* (2015) 'REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes', *Nucleic Acids Research*, 43(D1), pp. D298–D299. doi: 10.1093/nar/gku1046.

Samai, P. *et al.* (2015) 'Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity', *Cell*, 161(5), pp. 1164–1174. doi: 10.1016/j.cell.2015.04.027 [doi].

Sampson, T. R. *et al.* (2013) 'A CRISPR/Cas system mediates bacterial innate immune evasion and virulence', *Nature*, 497(7448), pp. 254–257. doi:

10.1038/nature12048.

Santos, A. P. *et al.* (2011) 'Genome of Mycoplasma haemofelis, unraveling its strategies for survival and persistence', *Veterinary Research*, 42(1). doi: 10.1186/1297-9716-42-102.

Schaefer, J. *et al.* (2016) 'Single-step method for β-galactosidase assays in Escherichia coli using a 96-well microplate reader', *Analytical Biochemistry*, 503, pp. 56–57. doi: 10.1016/j.ab.2016.03.017.

Schatz, D. G., Oettinger, M. A. and Baltimore, D. (1989) 'The V(D)J recombination activating gene, RAG-1', *Cell*, 59(6), pp. 1035–1048. doi: 0092-8674(89)90760-5 [pii].

Schouler, C. *et al.* (1998) 'Combinational variation of restriction modification specificities in Lactococcus lactis', *Molecular Microbiology*, 28(1), pp. 169–178. doi: 10.1046/j.1365-2958.1998.00787.x.

Schwarz, K. *et al.* (1996) 'RAG mutations in human B cell-negative SCID', *Science (New York, N.Y.)*, 274(5284), pp. 97–99.

Schwede, A. *et al.* (2015) 'How Does the VSG Coat of Bloodstream Form African Trypanosomes Interact with External Proteins?', *PLoS pathogens*, 11(12), p. e1005259. doi: 10.1371/journal.ppat.1005259 [doi].

Seo, G. J. *et al.* (2013) 'Reciprocal inhibition between intracellular antiviral signaling and the RNAi machinery in mammalian cells', *Cell host & microbe*, 14(4), pp. 435–445. doi: 10.1016/j.chom.2013.09.002 [doi].

Sitaraman, R., Denison, A. M. and Dybvig, K. (2002) 'A unique, bifunctional site-specific DNA recombinase from Mycoplasma pulmonis', *Molecular Microbiology*, 46, pp. 1033–1040.

Sitaraman, R. and Dybvig, K. (1997) 'The hsd loci of Mycoplasma pulmonis: organization, rearrangements and expression of genes.', *Molecular microbiology*, 26(1), pp. 109–120.

Sneppen, K. *et al.* (2015) 'Restriction modification systems as engines of diversity', *Frontiers in Microbiology*, 6(JUN), pp. 1–6. doi: 10.3389/fmicb.2015.00528.

Sniegowski, P. D., Gerrish, P. J. and Lenski, R. E. (1997) 'Evolution of high mutation rates in experimental populations of E-coli', *Nature*, 387(6634), pp. 703–705. doi: 10.1038/42701.

Sprouffske, K. and Wagner, A. (2016) 'Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves', 17. doi:

10.1186/s12859-016-1016-7.

Stern, A. *et al.* (2010) 'Self-targeting by CRISPR: gene regulation or autoimmunity?', *Trends in genetics : TIG*, 26(8), pp. 335–340. doi: 10.1016/j.tig.2010.05.008 [doi].

Stern, A. and Sorek, R. (2011) 'The phage-host arms race: Shaping the evolution of microbes', *BioEssays*, 33(1), pp. 43–51. doi: 10.1002/bies.201000071.

Sturrock, S. S. and Dryden, D. T. F. (1997) 'A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes', *Nucleic Acids Research*. Oxford University Press, 25(17), pp. 3408–3414. doi: 10.1093/nar/25.17.3408.

Su, X. Z. *et al.* (1995) 'The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes', *Cell*, 82(1), pp. 89–100. doi: 0092-8674(95)90055-1 [pii].

Swaney, S. *et al.* (1995) 'RNA-mediated resistance with nonstructural genes from the tobacco etch virus genome', *Molecular plant-microbe interactions : MPMI*, 8(6), pp. 1004–1011.

Swarts, D. C. *et al.* (2014) 'The evolutionary journey of Argonaute proteins', *Nature structural & molecular biology*, 21(9), pp. 743–753. doi: 10.1038/nsmb.2879 [doi].

Taddei, F. *et al.* (1997) 'Role of mutator alleles in adaptive evolution', *Nature*, 387(6634), pp. 700–702. doi: 10.1038/42696.

Tas, J. M. *et al.* (2016) 'Visualizing antibody affinity maturation in germinal centers', *Science (New York, N.Y.)*, 351(6277), pp. 1048–1054. doi: 10.1126/science.aad3439 [doi].

Teng, G. *et al.* (2015) 'RAG Represents a Widespread Threat to the Lymphocyte Genome', *Cell*, 162(4), pp. 751–765. doi: 10.1016/j.cell.2015.07.009 [doi].

Tettelin, H. *et al.* (2001) 'Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae', *Science*, 293(5529), pp. 498–506. doi: 10.1126/science.1061217.

Tubbs, A. and Nussenzweig, A. (2017) 'Endogenous DNA Damage as a Source of Genomic Instability in Cancer', *Cell*, 168(4), pp. 644–656. doi: S0092-8674(17)30005-3 [pii].

Vale, P. F. *et al.* (2015) 'Costs of CRISPR-Cas-mediated resistance in Streptococcus thermophilus', *Proceedings.Biological sciences*, 282(1812), p. 20151270. doi: 10.1098/rspb.2015.1270 [doi].

Vasu, K. and Nagaraja, V. (2013) 'Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense', *Microbiology and Molecular Biology Reviews*, 77(1), pp. 53–72. doi: 10.1128/MMBR.00044-12.

Vergara, D., Jokela, J. and Lively, C. M. (2014) 'Infection dynamics in coexisting sexual and asexual host populations: support for the Red Queen hypothesis', *The American Naturalist*, 184 Suppl, pp. S22-30. doi: 10.1086/676886 [doi].

Victora, G. D. *et al.* (2010) 'Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter', *Cell*, 143(4), pp. 592–605. doi: 10.1016/j.cell.2010.10.032 [doi].

de Visser, J. A. and Elena, S. F. (2007) 'The evolution of sex: empirical insights into the roles of epistasis and drift', *Nature reviews.Genetics*, 8(2), pp. 139–149. doi: nrg1985 [pii].

Vovis, G. F. and Zinder, N. D. (1975) 'Methylation of f1 DNA by a restriction endonuclease from Escherichia coli B', *Proceedings of the National Academy of Sciences of the United States of America*, 95(10), pp. 557–568. Available at: http://www.pnas.org/content/71/10/3810.full.pdf (Accessed: 11 September 2017).

Wang, X. H. *et al.* (2006) 'RNA interference directs innate immunity against viruses in adult Drosophila', *Science (New York, N.Y.)*, 312(5772), pp. 452–454. doi: 1125694 [pii].

Weinstein, J. A. *et al.* (2009) 'High-throughput sequencing of the zebrafish antibody repertoire', *Science (New York, N.Y.)*, 324(5928), pp. 807–810. doi: 10.1126/science.1170020 [doi].

Westra, E. R. *et al.* (2012) 'CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3', *Molecular Cell*, 46(5), pp. 595–605. doi: 10.1016/j.molcel.2012.03.018.

Westra, E. R. *et al.* (2015) 'Parasite exposure drives selective evolution of constitutive versus inducible defense', *Current Biology*, 25(8), pp. 1043–1049. doi: 10.1016/j.cub.2015.01.065.

Westra, E. R. *et al.* (2017) 'Mechanisms and consequences of diversity-

generating immune strategies', *Nature Reviews Immunology*. Nature Publishing Group. doi: 10.1038/nri.2017.78.

Westra, E. R., Buckling, A. and Fineran, P. C. (2014) 'CRISPR–Cas systems: beyond adaptive immunity', *Nature Reviews Microbiology*, 12(5), pp. 317–326. doi: 10.1038/nrmicro3241.

Wiedenheft, B. *et al.* (2011) 'RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions', *Proceedings of the National Academy of Sciences*, 108(25), pp. 10092–10097. doi: 10.1073/pnas.1102716108.

Wielgoss, S. *et al.* (2016) 'Adaptation to Parasites and Costs of Parasite Resistance in Mutator and Nonmutator Bacteria', *Molecular biology and evolution*, 33(3), pp. 770–782. doi: 10.1093/molbev/msv270 [doi].

Wilson, G. G. and Murray, N. E. (1991) 'Restriction and modification systems.', *Annual review of genetics*, 25, pp. 585–627. doi: 10.1146/annurev.genet.25.1.585.

Wolfe, K. H., Sharp, P. M. and Li, W. H. (1989) 'Mutation rates differ among regions of the mammalian genome', *Nature*, 337(6204), pp. 283–285. doi: 10.1038/337283a0 [doi].

Young, D., Hussell, T. and Dougan, G. (2002) 'Chronic bacterial infections: living with unwanted guests', *Nature immunology*, 3(11), pp. 1026–1032. doi: 10.1038/ni1102-1026 [doi].

Zaman, L. *et al.* (2014) 'Coevolution drives the emergence of complex traits and promotes evolvability', *PLoS biology*, 12(12), p. e1002023. doi: 10.1371/journal.pbio.1002023 [doi].

Zegans, M. E. *et al.* (2009) 'Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of Pseudomonas aeruginosa.', *Journal of bacteriology*. American Society for Microbiology, 191(1), pp. 210–9. doi: 10.1128/JB.00797-08.

Zhu, Y. *et al.* (2000) 'Genetic diversity and disease control in rice', *Nature*, 406(6797), pp. 718–722. doi: 10.1038/35021046 [doi].

Zinkevich, V. *et al.* (1992) 'Mutation in the specificity polypeptide of the type I restriction endonuclease R EcoK that affects subunit assembly', *Journal of Molecular Biology*, 227(3), pp. 597–601. doi: 10.1016/0022-2836(92)90210-B.