

The phage-host arms-race: Shaping the evolution of microbes

Adi Stern and Rotem Sorek*

Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100,
Israel

* Corresponding author: rotem.sorek@weizmann.ac.il

Abbreviations: CRISPR (clustered regularly interspersed palindromic repeats), LGT (lateral gene transfer), RM (restriction modification), MTase (methyltransferase), REase (restriction endonuclease), cas (CRISPR associated sequence) crRNA (CRISPR RNA), Abi (Abortive infection), toxin-antitoxin (TA).

Keywords: arms-race, phage, bacteria, evolution, resistance

Abstract

Bacteria, the most abundant organisms on the planet, are outnumbered by a factor of 10 to 1 by phages that infect them. Faced with the rapid evolution and turnover of phage particles, bacteria have evolved various mechanisms to evade phage infection and killing, leading to an evolutionary arms-race. The extensive co-evolution of both phage and host has resulted in considerable diversity on the part of both bacterial and phage defensive and offensive strategies. Here, we discuss the unique and common features of phage resistance mechanisms and their role in global biodiversity. The commonalities between defense mechanisms suggest avenues for the discovery of novel such mechanisms based on their evolutionary traits.

Introduction

Phage-host relationships have been studied intensively since the early days of molecular biology. In the late 1970s, while viruses were found to be ubiquitous, it was assumed that they were present in relatively low numbers and that their effect on microbial communities was low [1]. With the increasing availability of new molecular techniques that allow studies of microbial communities without the need to culture them [2], it is now realized that viruses greatly outnumber bacteria in the ocean and other environments, with viral numbers ($\sim 10^7$ – 10^8 ml⁻¹) often tenfold larger than bacterial cell counts ($\sim 10^6$ ml⁻¹) [3-5]. Thus, bacteria are confronted with a constant threat of phage predation.

The Red Queen hypothesis (Box 1) posits that competitive environmental interactions, such as those displayed by hosts and parasites, will lead to continuous variation and selection towards adaptation of the host, and counter-adaptations on the

side of the parasite. Arguably, nowhere is this evolutionary trend so pronounced as in phage-microbe interactions. This is due to the extremely rapid evolution and turnover of phage particles [6], causing acute pressure on microbial communities to evade infection and killing by phages. In fact, the arms-race between phage and bacteria is predicted to have had an impact on global nutrient cycling [7], on global climate [6, 7], on the evolution of the biosphere [8], and also on the evolution of virulence in human pathogens [9].

Box 1. The Red Queen Hypothesis

The Red Queen hypothesis was originally proposed by Leigh Van Valen (1973) [10], and is also termed the evolutionary arms-race hypothesis. As the Red Queen tells Alice in Lewis Carroll's "Through the Looking-Glass": "Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!" The original theory proposed that in tight co-evolutionary interactions such as those in a prey-predator relationship, changes (e.g., running faster) on the one side may lead to near extinction of the other side. The only way the second side can maintain its fitness is by counter-adaptation (running even faster). This will lead to an uneasy balance between prey and predator, where species have to constantly evolve in order to stay at the same fitness level. The metaphor of an evolutionary arms-race has been found to be relevant for many biological processes, but nowhere is this metaphor as apt as in host-parasite relationships.

This review focuses on the evolution of three of the most well studied microbial defense mechanisms against phage: the restriction-modification system, the recently discovered CRISPR (clustered regularly interspersed palindromic repeats) loci together with their associated *cas* genes, and the abortive infection system (summarized in Table 1). We first describe these defense systems, as well as the counter-adaptations that evolved in the phage to allow escape from bacterial defense. Next, we discuss features that are common to many microbial defense systems, such as rapid evolution, tendency for lateral gene transfer (LGT), and the selfish nature of these systems. We elaborate on defense systems which have gained new functions in the host genome. Finally, the exciting hypothesis that many other prokaryotic defense systems are still yet to be discovered is discussed. Our review mainly focuses on the evolutionary angle of the phage-host arms-race; for deeper mechanistic descriptions of phage resistance systems we refer the readers to an excellent recent review published elsewhere [11]. In addition, the discussion here is restricted to active defense mechanisms; passive host adaptations, such as mutations at the phage receptors, are not discussed here.

Restriction-modification systems: Degradation of foreign DNA

Arguably, the most well studied phage defense mechanism is the restriction-modification (RM) system [12], which is present in over 90% of sequenced bacterial and archaeal genomes [13]. This system is composed of two activities: one that restricts incoming foreign genetic material, and one that protects host genetic material from restriction (Fig. 1A). Both activities are mediated by recognition of a specific DNA sequence, on average 4-8 base-pairs (bp) long. Protection is normally conferred by modification (usually methylation) of specific bases in this recognition sequence in the host genome. Accordingly, all non-methylated DNA recognition sequences are recognized as foreign and are cleaved. Genetically, the minimal composition of RM systems consists of a methyltransferase (MTase) gene that perform the defense activity and a restriction endonuclease (REase) gene that performs the foreign restriction activity, and since the REases undergo rapid evolution, it is often the presence of a MTase that serves as the basis for identification of RM systems in newly sequenced genomes [14].

Phages have evolved to evade the ubiquitous RM systems in a variety of ways (Fig. 1B). Some phages have acquired an MTase, or stimulate the host MTase so that it confers protection to the phage genome [15]. Other phages code for proteins that target and shut down the REase. One interesting example is the Ocr protein of phage T7, which blocks the active site of some REases by mimicking 24 bp of bent B-form DNA [16]. Alternatively, some phages incorporate unusual bases in their genomes, thus throwing off the REase. For example, some *Bacillus subtilis* phages replace thymine with 5-hydroxymethyluracil [15]. These phages code a protein that then further inhibits the host protein uracil-DNA glycosylase from cleaving uracil bases

from the phage DNA [17, 18]. Interestingly, this inhibitory protein is also a DNA mimic (reviewed in [19]). The T-even phages T2, T4, and T6 also contain unusual bases in their genomes and may further post-synthetically glycosylate their DNA to avoid REase restriction [15]. Yet another evasion mechanism that has been reported is alteration of the restriction recognition sites. For example, some phages employ "palindrome avoidance": since Type II REases often recognize symmetrical (palindromic) sequences, some phages tend to avoid containing such sites in their genome [20].

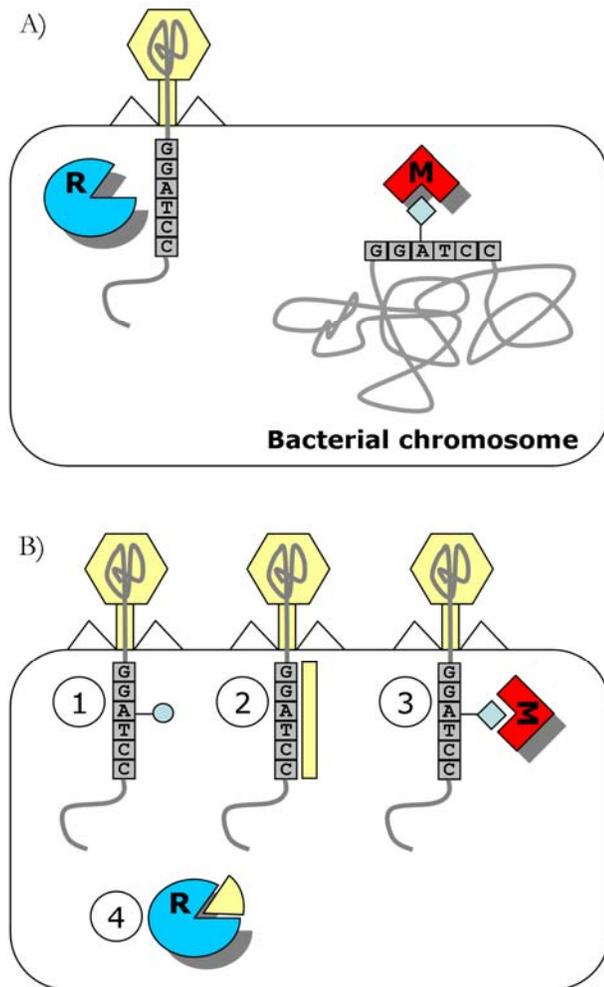


Figure 1. The restriction-modification defense system. A: A general illustration of function, exemplified by type II R-M enzymes. B: Examples of strategies employed by phage to evade restriction. (1) Incorporation of unusual bases protects from restriction [15]; (2) Masking of the restriction sites by phage proteins[112]; (3) Stimulation of MTase activity causes the phage DNA to be protected; (4) Neutralization of REase by phage proteins that mimic DNA [113].

The well-studied model organisms *Escherichia coli* and T4 phage display a fascinating example of a co-evolutionary arms-race [11]. The battle purportedly begins with the T4 phage genome, which contains the modified base hydroxymethylcytosine instead of cytosine [15]. To counter-attack the phage, *E. coli* K-12 possesses a unique form of REase, the McrBC enzyme, which cleaves only modified DNA substrates (such as that of T4) [21]. The T4 phage rises to the

challenge by glucosylating its genome, and is thus impervious to McrBC [15]. However, *E. coli* cT596 encodes the GmrS-GmrD system that can also restrict glucosylated DNA [22, 23]. In continuation of the battle, some T4 phages encode the IPI protein, which in its processed form (IPI*) disables the GmrS-GmrD system [24]. Evidently, the battle is far from an end, since bacterial strains have been found to overcome the IPI* protein of T4 [24]. In fact, it is likely that many bacterial defense systems, coupled with their cognate phage evasion strategies, have undergone similar attack and counter-attack cycles, where a change on one side selects for changes that can overcome the opponent.

CRISPR/Cas: Acquired adaptive immunity

The existence of a unique nucleotide arrangement in *E. coli*, comprised of a cluster of direct repeats interspersed with variable sequences (termed spacers; both around 30 bp long), was first described in 1987 [25]. These clusters, together with several CRISPR-associated (*cas*) genes (Fig. 2A), were later found to exist in ~40% of sequenced bacterial genomes and ~90% of archaeal genomes [26-28]. The first inkling that this system comprises a phage-defense mechanism arose when spacer sequences were found to be highly similar to DNA from foreign origin, i.e., from phage, plasmid or transposon DNA [29-31]. In 2007, experimental evidence was presented that this system indeed confers resistance to phage infection [32]. This led to the striking insight that similar to higher eukaryotes, bacteria and archaea possess acquired (and inherited) immunity [27, 33-35]. Following phage infection, it has been found that a small portion of bacterial cells integrated new spacers identical to the phage genomic sequence (termed proto-spacer), resulting in CRISPR-mediated phage resistance [32, 36, 37]. Further experiments have shown that the CRISPR locus is

transcribed into a single RNA transcript, which is then further cleaved by the Cas proteins to generate smaller CRISPR RNA (crRNA) units, each including one targeting spacer [37, 38]. These units then interfere with the incoming foreign genetic material by complementary base-pairing with either DNA [36, 37] or RNA [39] from the foreign element (Fig. 2A).

While the study of CRISPR/Cas is still in its infancy, examples of phages that are resistant to CRISPR/Cas interference have nevertheless been noted (Fig. 2B). Following the first round of infection and acquisition of novel spacers by the bacterial CRISPR, phages which have mutated, recombined or lost their proto-spacer target sequence in the second round of infection are now resistant to CRISPR [32, 40-42]. It also seems likely that phages have evolved mechanisms that directly target the CRISPR/Cas machinery. To date, only vague hints of such mechanisms exist. For example, it has been shown that one of the proteins encoded by the T7 phage phosphorylates the CasB protein [43]. It remains to be shown whether this feat of the phage affects CRISPR/Cas functioning.

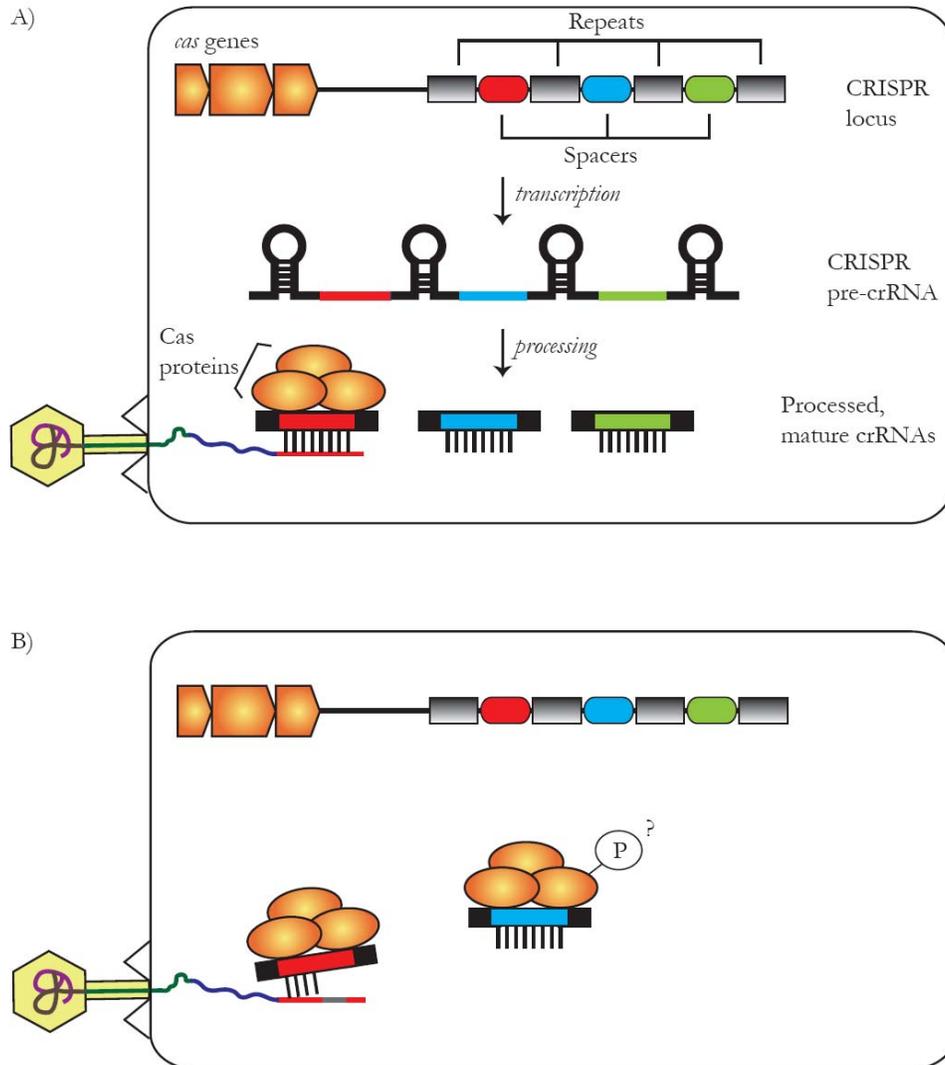


Figure 2. The CRISPR/Cas system. A: Mechanism of action: transcription from the repeat-spacer CRISPR locus generates a long non-coding RNA, with repeats that may sometimes assume a secondary structure. Cleavage of the repeat sequences by the Cas proteins generates crRNAs that target the phage DNA or RNA, and interfere with phage infection. B: Phages can evade CRISPR interference by mutation or recombination of the targeted proto-spacer sequence. Another putative evasion mechanism is phosphorylation of the Cas proteins. This remains, however, to be verified.

Abortive infection: Cellular suicide

If a phage has successfully entered the host cell and avoided restriction by the host

RM systems and by CRISPR, it proceeds to develop, replicate, and release its

progeny. Abortive infection (Abi) is a collective term describing host mechanisms that interrupt with phage development at different stages of phage transcription, genome replication, and phage packaging [44]. Abi-mediated resistance leads to death ("suicide") of the cell, and is thought to occur since corruption of host functions has already been initiated by the phage. However, this death confers an advantage to surrounding bacterial cells since it confines the infection to the sacrificed cell and prevents the spread of infectious particles.

Although several Abi systems have been discovered, the majority of them encoded on plasmids, the mechanism by which they operate remains largely unknown. Frequently, abortive infection is mediated by a single gene encoded on a plasmid or on a prophage, which displays little or no homology to any known proteins. Some Abi genes have been shown to target phage genes involved in DNA replication [45, 46], and others have been shown to target the host translation apparatus [47, 48]. For instance, in *E. coli* K-12 the Lit protease, encoded by the defective ϕ 14 prophage, is only activated in the presence of a short polypeptide called Gol, which is produced by the T4 phage. Once active, this protease cleaves the translation elongation factor Ef-Tu, thus leading to translational arrest and cell death (Fig. 3A). Similarly, the Prr protein (also part of a defective prophage in *E. coli* cT196) cleaves tRNA^{Lys} in response to the presence of the T4 peptide Stp [49]. Mutations in these phage peptides suppress activation of Lit or Prr and rescue the infecting phage [47, 49] (Fig. 3B).

Since Abi genes have a toxic effect on their host, they are under tight regulation [50]. In fact, many similarities can be drawn between Abi systems (and also RM systems), and toxin-antitoxin (TA) systems. TA systems are composed of a stable toxin and an unstable antitoxin [51]. Normally, the antitoxin binds and inhibits the

toxin. However, a decrease in the levels of the unstable antitoxin activates the toxin and leads to growth arrest or cell death. It has been shown that this situation occurs in response to phage infection in two known TA modules, *mazEF* and *hok-sok*, which can thus cause abortive infection [52, 53]. Moreover, the AbiQ system was recently found to function as a protein-RNA TA pair [54], albeit how exactly this system interferes with phage replication remains unknown. Thus, it appears that some TA systems may be a subtype of Abi systems. Interestingly, RM systems may also be viewed as TA systems, since loss of the MTase (which is often unstable, similar to the antitoxin) results in a toxic effect of the REase [55, 56].

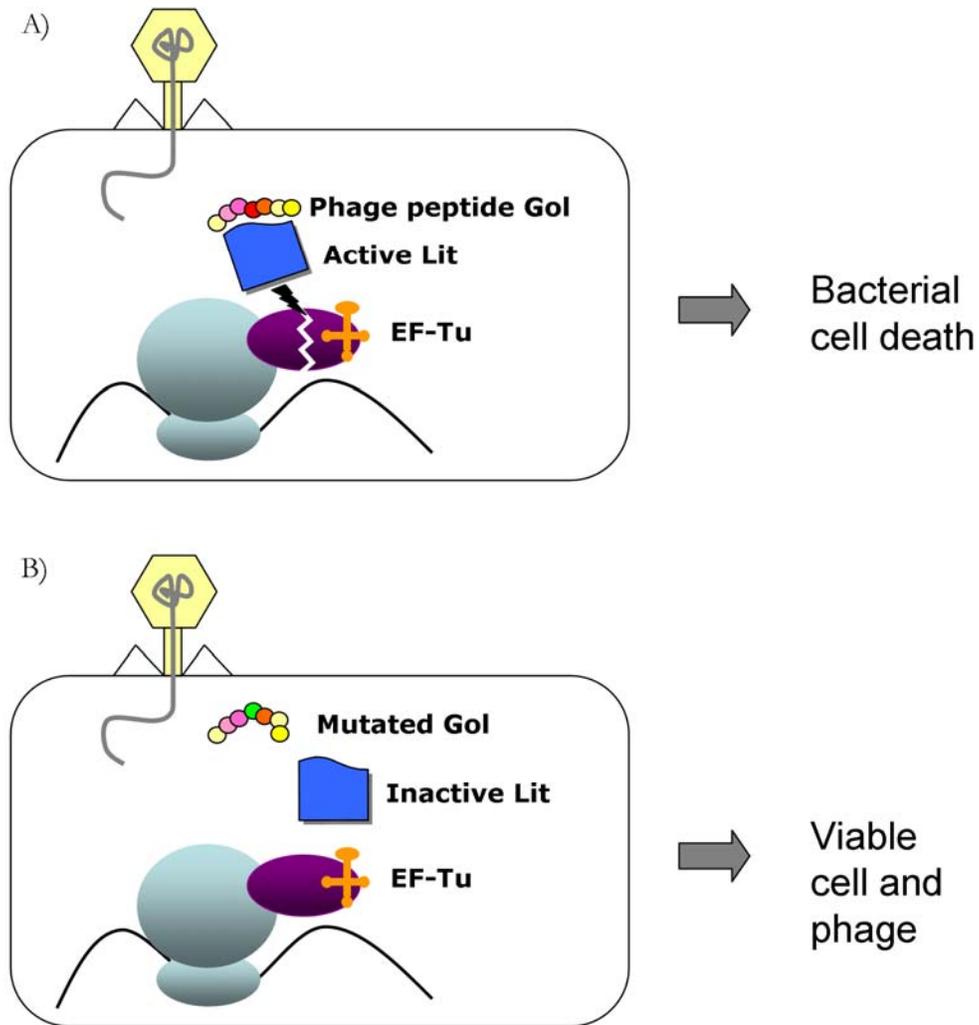


Figure 3. The abortive infection system. A: Illustration of the mechanism of the *E. coli* K-12 abortive infection Lit system. When the T4 phage peptide Gol is synthesized, it binds and activates the bacterial (prophage-encoded) Lit protein, which then cleaves the elongation factor EF-Tu. This leads to the arrest of protein synthesis and to bacterial cell death, with the phage trapped inside. B: A mutation at the Gol polypeptide reduces the activation of Lit, and rescues the phage.

Commonalities among phage defense systems

One of the most striking features common to all phage defense systems is their high genetic variability, which occurs as a consequence of the co-evolutionary arms-race with phages [57]. This is manifested in the enormous numbers and types of RM systems [12, 58], of CRISPR subtypes [59, 60], and in the variety of abortive

infection systems [44]. Moreover, the gene sequences of these systems often display a high evolutionary rate ([61-63], Stern and Sorek, unpublished data). This variability is most easily exemplified in the diversity of RM systems: Four inherently different groups of RM systems exist, classified as Type I-IV based on their subunit composition, mode of action, and cofactor requirement ([reviewed in 12]). Each type of RM system includes an assortment of restriction enzymes that recognize different recognition sequences. For instance, nearly 4000 Type II enzymes are known of today, which are further divided into 11 overlapping sub-classes [13, 64]. Interestingly, Type II REase sequences have frequently been found to be ORFan sequences, i.e., have no significant similarity to any other protein [65, 66]. Initially, this was seen as evidence for convergent evolution [67]. However, crystallographic studies have shown structural similarity among Type II REases, most likely representing a rapid evolutionary rate since their divergence from a common ancestor [68].

Another trait common to all defense systems presented here is their propensity to undergo LGT, sometimes between distantly related prokaryotes [69-72]. Often these systems reside on plasmids, on prophages, on genomic islands of foreign origin, or are linked to transposase genes. This mobility allows rapid acquisition and dissemination of new systems to counteract the invading phage, thus contributing to the extensive co-evolution with the prokaryotic parasites. However, there is an interesting alternative view, not necessarily mutually exclusive, which may explain the high variability and mobility of phage defense systems. Accordingly, these systems are selfish elements, as will be elaborated in the next section.

Defense systems: A burden or a benefit?

In this section, the advantages and disadvantages of microbial immune systems are reviewed. Such a discussion is incomplete without examining the pros and cons of the phage-host relationship. Evidently, for a single cell, phage predation and killing of the host are a strong disadvantage. However, phages may also contribute directly to host fitness by supplying a pool of new and possibly beneficial genes [73]. As such, phages serve as vessels of LGT, which is one of the most important forces in microbial evolution (Box 2).

What is the associated cost of encoding phage defense mechanisms? An obvious primary cost is the energy cost linked to carrying additional genetic cargo. The fact that often, only a small portion of a given bacterial population bears a plasmid or a prophage with a given defense mechanism may be seen as evidence for this cost. A second intriguing cost of encoding a phage defense mechanism is the risk of autoimmunity. Autoimmunity, classically defined in mammalian immune systems, is the failure of the immune system to recognize what is self and what is foreign, resulting in an immune response against self. In RM systems, this is a clear danger since the restriction enzyme is often more stable than the protecting methylase [74]. Abortive infection mechanisms are equally lethal, since errant function of the *Abi* gene will lead to cell death. Finally, the CRISPR system is also not "immune" to errors, and the frequent acquisition of self genetic material also leads to spacers that target the self-genome, potentially incurring autoimmunity [75] (see also [76]).

Paradoxically phages themselves often bear anti-phage defense systems, enabling their acquisition by host cells. To cite a few examples, the *HindIII* RM gene complex was found on a cryptic prophage in the *Haemophilus influenzae* genome [77], the *abiN* abortive infection gene is encoded on a prophage in *Lactococcus lactis*

subsp. *cremoris* S114, and even a CRISPR array was found within a *Clostridium difficile* prophage [78]. These results initially seem counterintuitive, since what benefit is there for the phage to carry such systems? One possible explanation is that this allows superinfection exclusion, thus preventing other phages from infecting an already infected cell [11, 78, 79]. However, this phenomenon has also been viewed as evidence for the selfish nature of phage defense mechanisms.

The behavior of defense mechanisms as selfish mobile elements has been extensively discussed for the case of RM systems [55, 80], but is also applicable for many Abi systems that operate as TA systems, which also have "selfish" properties [51, 81]. The main lines of evidence in favor of this view is that (a) RM systems destroy any other invading RM system, (b) any attempt to lose the RM system will result in the death of the host, and (c) RM systems are prone to extensive mobility, and are often associated with plasmids, phages, transposons, and integrons [55]. These characteristics of the RM systems lead to an increase of their relative frequency in the bacterial population. Thus, according to this hypothesis the defense incurred by RM systems on host cells is a mere by-product of the fact that RM systems defend themselves.

Box 2. Impact of defense systems on LGT

LGT is the process whereby genetic material is incorporated from a non-parental organism. There are three major mechanisms of LGT: (a) transduction by phages; (b) transformation of naked DNA; and (c) bacterial conjugation, which involves transfer of DNA via direct connections generated between a pair of cells. LGT is recognized as one of the most important mechanisms of genetic innovation in both bacteria and archaea [82-84]. In fact, it is now realized that much of the genomic diversity in prokaryotes is a consequence of LGT, rather than allelic differences at the same loci [85, 86].

However, when viewed from the point of view of a single cell, LGT has a negligible chance of contributing to the fitness of the organism. Evidently, transduction by lytic phages may severely compromise the survival of a bacterial colony. Plasmids and transposons may also decrease the fitness of an organism by integrating into crucial regions of the genomes, by adding an energetic cost involved with replication of excess DNA, and by expressing harmful genes [e.g., 87, 88, 89]. Finally, even if a new beneficial gene sequence is obtained, there is only a small probability that it will integrate into the current cellular network of an organism without causing deleterious effects [90, 91]. Thus, defense mechanisms such as those described here are crucial for maintaining the genetic identity of organisms, and protecting it against the constant bombardment of potentially detrimental foreign DNA.

Exaptations: Alternative functions of defense systems

Intriguingly, a number of anti-phage defense systems have evolved to gain a distinct function in cellular regulation that is independent of phage restriction. Such evolutionary events were coined "exaptations" [92], a term used to describe the use of a biological structure or function for a purpose other than that for which it initially evolved. For instance, several RM systems have lost their REase activity, leaving an orphan MTase that can now take part in epigenetic modifications. The Dam methylase in *E. coli* and the CcrM methylase in *Caulobacter crescentus*, both of which have originated from RM systems [93], are two such examples. Methylation by Dam has been linked to several important regulatory processes such as mismatch repair by the MutHLS complex, binding of the replication initiation complex to methylated *OriC*, and regulation of bacterial pathogenicity [93]. On the other hand, the CcrM methylase has been shown to affect the cell cycle in alpha proteobacteria that encode this gene [93].

Another interesting example of exaptation of RM systems is evident in phase-variable Type III RM systems [94], whose genes can be reversibly inactivated due to tandem repeat tracts in their sequences. These repeats initiate a mechanism called slipped-strand mispairing, leading to a change in the number of repeats after DNA replication and possible frame-shift mutations [94]. Several regulatory roles have been suggested for phase-variable RM systems: (a) to allow regulated removal of the barrier against foreign DNA, thus allowing potentially beneficial uptake of DNA [95, 96], (b) autolytic self-DNA degradation, or "bacterial suicide" [97-99], further discussed below, and (c) epigenetic gene regulation via differential methylation of the genome [100]. The latter phenomenon, which allows switching different genes on and

off, has been linked to pathogenicity of bacterial species by allowing colonization, immune evasion and adaptation to novel environments [94].

TA modules, which include some abortive infection systems, are also known to participate in a variety of other cellular processes. For example, one of the most studied TA loci *mazEF*, aborts translation by cleaving mRNA molecules in response to different stress signals [101, 102], one of which is phage infection [52]. An ongoing debate exists whether this action is reversible or not: reversible effects have been attributed to bacteriostatic effects, which allow reduced growth rate of each cell during nutritional stress [101, 102]. On the other hand, irreversible effects of *mazEF* have been attributed to programmed cell death that occurs in a subpopulation of cells, permitting the survival of the population as a whole [103]. Interestingly, in *Myxococcus xanthus* it was shown that the toxin MazF exists without the antitoxin MazE, and has adopted a key transcriptional regulator as an alternative antitoxin [104]. MazF mediates programmed cell death during multicellular development of this organism. To summarize, these accounts exemplify the broad evolutionary diversification of different microbial defense mechanisms, and their potential to cross boundaries from phage-encoded mechanisms, to anti-phage systems, to regulatory host mechanisms.

Conclusions

Despite our growing understanding of microbial immunity, much still remains obscure. Do defense systems work separately or in unison? What is the cost of each system? Which phages are targeted by which systems? For instance, while most characterized defense systems work against double-stranded DNA phages, RNA

viruses might also be abundant [105]. Defense systems that target such viruses are yet to be discovered.

The recently discovered CRISPR system epitomizes our incomplete understanding of the complexity of bacterial defense systems. The discovery that almost half of all prokaryotes possess acquired inherited immunity came as a surprise to the scientific community, given our initial tendency to view prokaryotes as less complex organisms. However, since it is now realized that prokaryotes are faced with a constant threat of predation, it is becoming clearer that the microbial immune system must be highly complex. Nevertheless, the CRISPR system and the Abi systems have a highly sporadic distribution across the prokaryotic phylogeny (Table 1). Furthermore, while RM systems are present in ~90% of prokaryotic species [13], this system is far from tight-proof, since it has been shown that phages have a non-negligible probability of escaping the REase and being methylated by the MTase. [106]. Thus, all three mechanisms reviewed here are either somewhat leaky, or span only part of the phylogenetic breadth of bacteria and archaea. Given the intensive arms-race between bacteria and phage, it seems probable that there are yet other unknown defense mechanisms waiting to be revealed.

So how does one set about the search for novel prokaryotic defense systems, and how does one learn more about existing ones? It is likely that many solutions to these questions will be made possible thanks to advances in sequencing technologies, such as whole-transcriptome studies and metagenomics. For instance, a seminal work by Andersson and Banfield [42] reconstructed viral and host population genomes from community genomic data, and showed how CRISPR spacers correlate with location of coexisting viruses and hosts. In addition, controlled experimental evolution studies, such as those performed in *Pseudomonas fluorescens* and its phage,

enable a direct survey of the changes that occur in the phage and host population with and without infection [107-109]. These studies are also likely to shed light on current defense systems and possibly discover novel ones.

More specifically, the search for novel defense mechanisms may be based on the well established common properties of both eukaryotic and prokaryotic immune systems: their high rate of evolution, and their tendency to undergo horizontal gene transfer. Thus, it may be that a reservoir of novel defense mechanisms lies in the most variable regions of bacterial and archaeal genomes, known as genomic islands [110]. A strategy that focuses on such islands to search for novel phage resistance mechanisms might lead, in the future, to surprising discoveries.

Disclaimer

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Acknowledgements

R.S. is an EMBO Young Investigator. He was supported, in part, by the ISF-FIRST program (grant 1615/09), NIH R01AI082376-01, ERC-StG, the Wolfson Family Trust miRNA research program, the Minerva foundation, and the Yeda-Sela Center for basic research. A.S. was supported by the Clore postdoctoral fellowship. This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

1. **Torrella, F. and R.Y. Morita**, Evidence by electron micrographs for a high incidence of bacteriophage particles in the waters of Yaquina Bay, Oregon:

- ecological and taxonomical implications. *Appl Environ Microbiol*, 1979. **37**: 774-8.
2. **Riesenfeld, C.S., P.D. Schloss, and J. Handelsman**, Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet*, 2004. **38**: 525-52.
 3. **Bergh, O., et al.**, High abundance of viruses found in aquatic environments. *Nature*, 1989. **340**: 467-8.
 4. **Chibani-Chennoufi, S., et al.**, Phage-host interaction: an ecological perspective. *Journal of bacteriology*, 2004. **186**: 3677.
 5. **Wommack, K.E. and R.R. Colwell**, Virioplankton: viruses in aquatic ecosystems. *Microbiology and Molecular Biology Reviews*, 2000. **64**: 69.
 6. **Fuhrman, J.A.**, Marine viruses and their biogeochemical and ecological effects. *Nature*, 1999. **399**: 541-548.
 7. **Suttle, C.A.**, Marine viruses—major players in the global ecosystem. *Nature Reviews Microbiology*, 2007. **5**: 801-812.
 8. **Comeau, A.M. and H.M. Krisch**, War is peace--dispatches from the bacterial and phage killing fields. *Current opinion in microbiology*, 2005. **8**: 488-494.
 9. **Brussow, H., C. Canchaya, and W.D. Hardt**, Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev*, 2004. **68**: 560-602.
 10. **Van Valen, L.**, A new evolutionary law. *Evolutionary theory*, 1973. **1**: 1-30.
 11. **Labrie, S.J., J.E. Samson, and S. Moineau**, Bacteriophage resistance mechanisms. *Nat Rev Microbiol*, 2010. **8**: 317-327.
 12. **Toock, M.R. and D.T. Dryden**, The biology of restriction and anti-restriction. *Curr Opin Microbiol*, 2005. **8**: 466-72.
 13. **Roberts, R.J., et al.**, REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res*, 2010. **38**: D234-6.
 14. **Roberts, R.J., et al.**, REBASE--restriction enzymes and DNA methyltransferases. *Nucleic acids research*, 2005. **33**: D230.
 15. **Kruger, D.H. and T.A. Bickle**, Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol Rev*, 1983. **47**: 345-60.
 16. **Bandyopadhyay, P.K., et al.**, Inhibition of the type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. *J Mol Biol*, 1985. **182**: 567-78.
 17. **Wang, Z. and D.W. Mosbaugh**, Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. *J Biol Chem*, 1989. **264**: 1163-71.
 18. **Wang, Z. and D.W. Mosbaugh**, Uracil-DNA glycosylase inhibitor of bacteriophage PBS2: cloning and effects of expression of the inhibitor gene in *Escherichia coli*. *J Bacteriol*, 1988. **170**: 1082-91.
 19. **Putnam, C.D. and J.A. Tainer**, Protein mimicry of DNA and pathway regulation. *DNA Repair (Amst)*, 2005. **4**: 1410-20.
 20. **Rocha, E.P., A. Danchin, and A. Viari**, Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. *Genome Res*, 2001. **11**: 946-58.
 21. **Raleigh, E.A. and G. Wilson**, *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. *Proc Natl Acad Sci U S A*, 1986. **83**: 9070-4.

22. **Bair, C.L., D. Rifat, and L.W. Black**, Exclusion of glucosyl-hydroxymethylcytosine DNA containing bacteriophages is overcome by the injected protein inhibitor IPI*. *J Mol Biol*, 2007. **366**: 779-89.
23. **Bair, C.L. and L.W. Black**, A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *J Mol Biol*, 2007. **366**: 768-78.
24. **Rifat, D., et al.**, Restriction endonuclease inhibitor IPI* of bacteriophage T4: a novel structure for a dedicated target. *J Mol Biol*, 2008. **375**: 720-34.
25. **Ishino, Y., et al.**, Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology*, 1987. **169**: 5429.
26. **Jansen, R., et al.**, Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology*, 2002. **43**: 1565-1575.
27. **Sorek, R., V. Kunin, and P. Hugenholtz**, CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol*, 2008. **6**: 181-6.
28. **Grissa, I., G. Vergnaud, and C. Pourcel**, The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, 2007. **8**: 172.
29. **Bolotin, A., et al.**, Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 2005. **151**: 2551-61.
30. **Mojica, F.J., et al.**, Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*, 2005. **60**: 174-82.
31. **Pourcel, C., G. Salvignol, and G. Vergnaud**, CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 2005. **151**: 653-63.
32. **Barrangou, R., et al.**, CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 2007. **315**: 1709-12.
33. **Horvath, P. and R. Barrangou**, CRISPR/Cas, the immune system of bacteria and archaea. *Science*, 2010. **327**: 167-70.
34. **Marraffini, L.A. and E.J. Sontheimer**, CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet*, 2010. **11**: 181-190.
35. **van der Oost, J., et al.**, CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci*, 2009. **34**: 401-7.
36. **Marraffini, L.A. and E.J. Sontheimer**, CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, 2008. **322**: 1843.
37. **Brouns, S.J., et al.**, Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 2008. **321**: 960-4.
38. **Makarova, K.S., et al.**, A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*, 2006. **1**: 7.
39. **Hale, C.R., et al.**, RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, 2009. **139**: 945-956.
40. **Deveau, H., et al.**, Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J Bacteriol*, 2008. **190**: 1390-400.

41. **Heidelberg, J.F., et al.**, Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. *PLoS One*, 2009. **4**: e4169.
42. **Andersson, A.F. and J.F. Banfield**, Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*, 2008. **320**: 1047-50.
43. **Qimron, U., S. Tabor, and C.C. Richardson**, New Details about Bacteriophage T7-Host Interactions. *Microbe*, 2010. **5**: 117-120.
44. **Chopin, M.C., A. Chopin, and E. Bidnenko**, Phage abortive infection in lactococci: variations on a theme. *Current opinion in microbiology*, 2005. **8**: 473-479.
45. **Durmaz, E. and T.R. Klaenhammer**, Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected *Lactococcus lactis*. *Journal of bacteriology*, 2007. **189**: 1417.
46. **Bidnenko, E., D. Ehrlich, and M.C. Chopin**, Phage operon involved in sensitivity to the *Lactococcus lactis* abortive infection mechanism AbiD1. *Journal of bacteriology*, 1995. **177**: 3824.
47. **Georgiou, T., et al.**, Specific peptide-activated proteolytic cleavage of *Escherichia coli* elongation factor Tu. *Proceedings of the National Academy of Sciences of the United States of America*, 1998. **95**: 2891.
48. **Morad, I., et al.**, Functional expression and properties of the tRNA (Lys)-specific core anticodon nuclease encoded by *Escherichia coli* prrC. *Journal of Biological Chemistry*, 1993. **268**: 26842.
49. **Snyder, L.**, Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol Microbiol*, 1995. **15**: 415-20.
50. **Blower, T.R., et al.**, Mutagenesis and functional characterization of the RNA and protein components of the toxIN abortive infection and toxin-antitoxin locus of *Erwinia*. *J Bacteriol*, 2009. **191**: 6029-39.
51. **Gerdes, K., S.K. Christensen, and A. Lobner-Olesen**, Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol*, 2005. **3**: 371-82.
52. **Hazan, R. and H. Engelberg-Kulka**, *Escherichia coli* mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. *Molecular Genetics and Genomics*, 2004. **272**: 227-234.
53. **Pecota, D.C. and T.K. Wood**, Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. *Journal of bacteriology*, 1996. **178**: 2044.
54. **Fineran, P.C., et al.**, The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci U S A*, 2009. **106**: 894-9.
55. **Kobayashi, I.**, Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res*, 2001. **29**: 3742-56.
56. **Hayes, F.**, Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science*, 2003. **301**: 1496-9.
57. **Hoskisson, P.A. and M.C. Smith**, Hypervariation and phase variation in the bacteriophage 'resistome'. *Curr Opin Microbiol*, 2007. **10**: 396-400.
58. **Orlowski, J. and J.M. Bujnicki**, Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Research*, 2008. **36**: 3552.

59. **Haft, D.H., et al.**, A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol*, 2005. **1**: e60.
60. **Kunin, V., R. Sorek, and P. Hugenholtz**, Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol*, 2007. **8**: R61.
61. **Sharp, P.M., et al.**, Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. *Proc Natl Acad Sci U S A*, 1992. **89**: 9836-40.
62. **Murray, N.E., et al.**, Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. *Mol Microbiol*, 1993. **9**: 133-43.
63. **Zheng, Y., R.J. Roberts, and S. Kasif**, Identification of genes with fast-evolving regions in microbial genomes. *Nucleic Acids Res*, 2004. **32**: 6347-57.
64. **Roberts, R.J., et al.**, A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res*, 2003. **31**: 1805-12.
65. **Kroger, M., et al.**, Eight new restriction endonucleases from *Herpetosiphon giganteus*-divergent evolution in a family of enzymes. *Nucleic acids research*, 1984. **12**: 3127.
66. **Mullings, R., S.P. Bennett, and N.L. Brown**, Investigation of sequence homology in a group of type-II restriction/modification isoschizomers. *Gene*, 1988. **74**: 245-251.
67. **Wilson, G.G. and N.E. Murray**, Restriction and modification systems. *Annual review of genetics*, 1991. **25**: 585-627.
68. **Venclovas, C., A. Timinskas, and V. Siksnys**, Five-stranded beta-sheet sandwiched with two alpha-helices: a structural link between restriction endonucleases EcoRI and EcoRV. *Proteins*, 1994. **20**: 279-82.
69. **Nelson, K.E., et al.**, Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature*, 1999. **399**: 323-9.
70. **Jeltsch, A. and A. Pingoud**, Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J Mol Evol*, 1996. **42**: 91-6.
71. **Haaber, J., S. Moineau, and K. Hammer**, Activation and transfer of the chromosomal phage resistance mechanism *AbiV* in *Lactococcus lactis*. *Appl Environ Microbiol*, 2009. **75**: 3358-61.
72. **Godde, J.S. and A. Bickerton**, The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol*, 2006. **62**: 718-29.
73. **Gogarten, J.P. and J.P. Townsend**, Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol*, 2005. **3**: 679-87.
74. **Naito, T., K. Kusano, and I. Kobayashi**, Selfish behavior of restriction-modification systems. *Science*, 1995. **267**: 897-9.
75. **Stern, A., et al.**, Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet*, 2010. **26**: 335-40.
76. **Marraffini, L.A. and E.J. Sonthheimer**, Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, 2010. **463**: 568-71.

77. **Hendrix, R.W., et al.**, Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A*, 1999. **96**: 2192-7.
78. **Sebahia, M., et al.**, The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet*, 2006. **38**: 779-86.
79. **Lu, M.J. and U. Henning**, Superinfection exclusion by T-even-type coliphages. *Trends Microbiol*, 1994. **2**: 137-9.
80. **Nakayama, Y. and I. Kobayashi**, Restriction-modification gene complexes as selfish gene entities: roles of a regulatory system in their establishment, maintenance, and apoptotic mutual exclusion. *Proc Natl Acad Sci U S A*, 1998. **95**: 6442-7.
81. **Makarova, K.S., Y.I. Wolf, and E.V. Koonin**, Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol Direct*, 2009. **4**: 19.
82. **Gogarten, J.P., W.F. Doolittle, and J.G. Lawrence**, Prokaryotic evolution in light of gene transfer. *Mol Biol Evol*, 2002. **19**: 2226-38.
83. **Koonin, E.V., K.S. Makarova, and L. Aravind**, Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol*, 2001. **55**: 709-42.
84. **Dagan, T., Y. Artzy-Randrup, and W. Martin**, Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci U S A*, 2008. **105**: 10039-44.
85. **Whitaker, R.J. and J.F. Banfield**, Population genomics in natural microbial communities. *Trends Ecol Evol*, 2006. **21**: 508-16.
86. **Medini, D., et al.**, Microbiology in the post-genomic era. *Nat Rev Microbiol*, 2008. **6**: 419-30.
87. **Modi, R.I. and J. Adams**, Coevolution in bacterial-plasmid populations. *Evolution*, 1991. **45**: 656-667.
88. **Zund, P. and G. Lebek**, Generation time-prolonging R plasmids: correlation between increases in the generation time of *Escherichia coli* caused by R plasmids and their molecular size. *Plasmid*, 1980. **3**: 65-9.
89. **Bouma, J.E. and R.E. Lenski**, Evolution of a bacteria/plasmid association. *Nature*, 1988. **335**: 351-2.
90. **Kedzierska, B., et al.**, Toxicity of the bacteriophage lambda cII gene product to *Escherichia coli* arises from inhibition of host cell DNA replication. *Virology*, 2003. **313**: 622-8.
91. **Lercher, M.J. and C. Pal**, Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol Biol Evol*, 2008. **25**: 559-67.
92. **Brosius, J. and S.J. Gould**, On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". *Proc Natl Acad Sci U S A*, 1992. **89**: 10706-10.
93. **Marinus, M.G. and J. Casades**, Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev*, 2009. **33**: 488-503.
94. **Srikhanta, Y.N., et al.**, The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *Proc Natl Acad Sci U S A*, 2005. **102**: 5547-51.

95. **Ando, T., et al.**, Restriction-modification system differences in *Helicobacter pylori* are a barrier to interstrain plasmid transfer. *Mol Microbiol*, 2000. **37**: 1052-65.
96. **Donahue, J.P., et al.**, Overcoming the restriction barrier to plasmid transformation of *Helicobacter pylori*. *Mol Microbiol*, 2000. **37**: 1066-74.
97. **Dybvig, K., R. Sitaraman, and C.T. French**, A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proc Natl Acad Sci U S A*, 1998. **95**: 13923-8.
98. **Saunders, N.J., et al.**, Simple sequence repeats in the *Helicobacter pylori* genome. *Mol Microbiol*, 1998. **27**: 1091-8.
99. **Hamilton, H.L. and J.P. Dillard**, Natural transformation of *Neisseria gonorrhoeae*: from DNA donation to homologous recombination. *Mol Microbiol*, 2006. **59**: 376-85.
100. **Seib, K.L., I.R. Peak, and M.P. Jennings**, Phase variable restriction-modification systems in *Moraxella catarrhalis*. *FEMS Immunol Med Microbiol*, 2002. **32**: 159-65.
101. **Pedersen, K., S.K. Christensen, and K. Gerdes**, Rapid induction and reversal of a bacteriostatic condition by controlled expression of toxins and antitoxins. *Mol Microbiol*, 2002. **45**: 501-10.
102. **Christensen, S.K., et al.**, Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. *J Mol Biol*, 2003. **332**: 809-19.
103. **Engelberg-Kulka, H., et al.**, Bacterial programmed cell death and multicellular behavior in bacteria. *PLoS Genet*, 2006. **2**: e135.
104. **Nariya, H. and M. Inouye**, MazF, an mRNA interferase, mediates programmed cell death during multicellular *Myxococcus* development. *Cell*, 2008. **132**: 55-66.
105. **Lang, A.S., et al.**, RNA viruses in the sea. *FEMS Microbiol Rev*, 2009. **33**: 295-323.
106. **Korona, R. and B.R. Levin**, Phage-mediated selection and the evolution and maintenance of restriction-modification. *Evolution*, 1993. **47**: 556-575.
107. **Paterson, S., et al.**, Antagonistic coevolution accelerates molecular evolution. *Nature*, 2010. **464**: 275-8.
108. **Buckling, A., et al.**, Antagonistic coevolution with parasites increases the cost of host deleterious mutations. *Proc Biol Sci*, 2006. **273**: 45-9.
109. **Wichman, H.A., et al.**, Different trajectories of parallel evolution during viral adaptation. *Science*, 1999. **285**: 422-4.
110. **Juhas, M., et al.**, Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev*, 2009. **33**: 376-93.
111. **Pandey, D.P. and K. Gerdes**, Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res*, 2005. **33**: 966-76.
112. **Iida, S., et al.**, Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA- phages. *Virology*, 1987. **157**: 156-66.
113. **Dryden, D.T. and M.R. Tock**, DNA mimicry by proteins. *Biochem Soc Trans*, 2006. **34**: 317-9.

Table 1. A summary of the major defense mechanisms described in the manuscript.

Name	Mechanism	Phylogenetic breadth
Restriction-modification	The restriction enzyme cleaves specific patterns in the incoming foreign DNA, while the modification enzyme protects host DNA from cleavage by unique biochemical modification.	Appear in ~90% of all sequenced prokaryotic genomes [13]
CRISPR/Cas	Fragments of phage DNA are integrated into CRISPR loci, which are then transcribed and processed into short non-coding RNAs. These RNAs, along with the associated Cas proteins, guide the way to interfere with the phage nucleic acids, in a yet unknown mechanism.	Appear in ~40% and ~90% of all sequenced bacterial and archaeal genomes, respectively [27]. The distribution of CRISPR/Cas-bearing species across the phylogenetic tree of life is highly patchy [38, 59].
Abortive infection	Premature cellular death occurs upon phage entry, blocking the expansion of the phage to neighboring cells. Notably, abortive infection systems include a large collection of mechanisms with little or no known evolutionary relationship, apart from a very similar phenotype.	Currently known abortive infection systems display a sporadic phylogenetic distribution in Gamma-proteobacteria and in Firmicutes [44], yet recent reports suggest that some systems may have an even broader phylogenetic range [54, 111].