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

Adi Stern and Rotem Sorek*

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

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 (~10^9/mL) often 10-fold larger than bacterial cell counts (~10^6/mL) [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 phages 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].

This review focuses on the evolution of three of the most well-studied microbial defense mechanisms against phages: the restriction-modification (RM) system, the recently discovered “clustered regularly interspersed palindromic repeats” (CRISPR) loci together with CRISPR-associated (cas) genes, and the abortive infection (Abi) system (summarized in Table 1). We first describe these defense systems, and 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 that have gained new functions in the host genome. Finally, the exciting hypothesis that many other prokaryotic defense systems are still to be discovered is discussed. Our review mainly focuses on the evolutionary angle of the phage-
Box 1

The Red Queen hypothesis

The Red Queen hypothesis was originally proposed by Van Valen [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 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.

Restriction-modification systems: Degradation of foreign DNA

Arguably, the most well-studied phage defense mechanism is the 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 another that protects host genetic material from restriction (Fig. 1A). Both activities are mediated by recognition of a specific DNA sequence, on average 4–8 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 performs the defense activity and a restriction endonuclease (REase) gene that performs the foreign restriction activity. Since the REases undergo rapid evolution, it is often the presence of an 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 ref. 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 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].

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 thymine.

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

<table>
<thead>
<tr>
<th>Name</th>
<th>Mechanism</th>
<th>Phylogenetic breadth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restriction-modification</td>
<td>The restriction enzyme cleaves specific patterns in the incoming foreign DNA, while the modification enzyme protects host DNA from cleavage by unique biochemical modification</td>
<td>Appear in ~90% of all sequenced prokaryotic genomes [13]</td>
</tr>
<tr>
<td>CRISPR/Cas</td>
<td>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 an as-yet-unknown mechanism</td>
<td>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]</td>
</tr>
<tr>
<td>Abortive infection</td>
<td>Premature cellular death occurs upon phage entry, blocking the expansion of the phage to neighboring cells. Notably, Abi systems include a large collection of mechanisms with little or no known evolutionary relationship, apart from a very similar phenotype</td>
<td>Currently known Abi 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]</td>
</tr>
</tbody>
</table>
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/C3) disables the GmrS-GmrD system [24]. Evidently, the battle is far from an end, since bacterial strains have been found to overcome the IPI/C3 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 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 vertebrates, prokaryotes 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 that have mutated, recombined or lost their proto-spacer target sequence in the second round of infection are then 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.

**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. 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, with the majority of them encoded on plasmids, the mechanism by which they operate remains largely unknown. Frequently, Abi 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 e14 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 Abi [52, 53]. Moreover, the AbiQ system was recently found to function as a protein-RNA TA pair [54], albeit exactly how 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 like the antitoxin is often unstable) results in a toxic effect of the REase [55, 56].

**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 Abi systems [44]. Moreover, the gene sequences of these systems often display a high evolutionary rate ([61–63]; A. Stern and R. 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 ref. 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 subclasses [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, or 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 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]. Abi 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 ref. 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 Abi 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.

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 REnase 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 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 [96]. 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] (discussed further 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 [96].

TA modules, which include some Abi 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 of 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 phylony (Table 1). Furthermore, while RM systems are present in ~90% of prokaryotic species [13], this system is far from infallible, 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 phages, it seems probable that there are other yet 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 the 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 LGT. 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.

Acknowledgments
R.S. is an EMBO Young Investigator. He was supported, in part, by the ISF-FIRST program (grant 1615/09), NIH R01AI082376-01, ERC-SIG, 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 post-doctoral fellowship.

References


