Phages and their satellites encode hotspots of antiviral systems

Graphical abstract

Highlights

- P2 phages and P4 satellites carry a genetic hotspot of diverse anti-phage systems
- The PARIS system kills the host in the presence of a phage anti-restriction protein
- Satellites are parasites of phages, but anti-phage systems can make them mutualistic
- Hotspots of anti-phage systems are found in prophages of diverse species

Authors
François Rousset, Florence Depardieu, Solange Miele, ..., Eduardo P.C. Rocha, Aude Bernheim, David Bikard

Correspondence
rousset.fra@gmail.com (F.R.), david.bikard@pasteur.fr (D.B.)

In brief
Rousset, Depardieu et al. show that E. coli P2-like phages and their parasitic P4-like satellites carry diverse anti-phage systems. The PARIS system induces abortive infection, triggered by a phage-encoded anti-restriction protein. Anti-phage systems can turn the parasitic satellite-phage relationship into a mutualistic one when facing a common threat.
Phages and their satellites encode hotspots of antiviral systems

François Rousset, 1,5,6,* Florence Depardieu, 1,5 Solange Miele, 1 Julien Dowding, 1 Anne-Laure Laval, 1 Erica Lieberman, 2 Daniel Garry, 2 Eduardo P.C. Rocha, 3 Aude Bernheim, 4 and David Bikard 1,7, *

1Institut Pasteur, Université de Paris, CNRS UMR 6047, Synthetic Biology, 75015 Paris, France
2Eligo Bioscience, Paris, France
3Institut Pasteur, Université de Paris, CNRS UMR 3525, Microbial Evolutionary Genomics, 75015 Paris, France
4Université de Paris, INSERM, IAME, 75006 Paris, France
5These authors contributed equally
6Present address: Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel
7Lead contact
*Correspondence: rousset.fra@gmail.com (F.R.), david.bikard@pasteur.fr (D.B.)
https://doi.org/10.1016/j.chom.2022.02.018

SUMMARY

Bacteria carry diverse genetic systems to defend against viral infection, some of which are found within prophages where they inhibit competing viruses. Phage satellites pose additional pressures on phages by hijacking key viral elements to their own benefit. Here, we show that E. coli P2-like phages and their parasitic P4-like satellites carry hotspots of genetic variation containing reservoirs of anti-phage systems. We validate the activity of diverse systems and describe PARIS, an abortive infection system triggered by a phage-encoded anti-restriction protein. Antiviral hotspots participate in inter-viral competition and shape dynamics between the bacterial host, P2-like phages, and P4-like satellites. Notably, the anti-phage activity of satellites can benefit the helper phage during competition with virulent phages, turning a parasitic relationship into a mutualistic one. Anti-phage hotspots are present across distant species and constitute a substantial source of systems that participate in the competition between mobile genetic elements.

INTRODUCTION

Bacteria employ an arsenal of defense strategies to overcome infection by bacteriophages (Bernheim and Sorek, 2020). Recent studies have shown that bacterial immunity is much more diverse that previously envisioned, spanning various mechanisms including DNA restriction (Goldfarb et al., 2015; Kuzmenko et al., 2020; Ofir et al., 2018, 2021; Xiong et al., 2020), abortive infection (Bobonis et al., 2020a; Cohen et al., 2019; Depardieu et al., 2016; Doron et al., 2018; Dy et al., 2014; Lopatina et al., 2020; Millman et al., 2020a; Owen et al., 2021), chemical interference (Bernheim et al., 2021; Kronheim et al., 2018), or nucleotide depletion (Tal et al., 2021a). A series of remarkable discoveries stemmed from the observation that bacterial defense systems tend to cluster in genomic regions termed defense islands (Makarova et al., 2011, 2013). The systematic investigation of genes found in association with known defense genes has considerably expanded our knowledge of bacterial immunity (Doron et al., 2018; Gao et al., 2020). Despite these notable findings, it is believed that many defense systems remain to be discovered. Beyond the description of the defensive arsenal of bacteria, the mechanisms that drive the variability, diversity, and number of defensive functions in bacterial genomes are just beginning to be explored (Tesson et al., 2021).

Recent studies of large ecological datasets of phage-bacteria interactions in Vibrio have highlighted how mobile genetic elements (MGEs) such as the SXT integrative conjugative element (ICE) carry anti-phage functions (LeGault et al., 2021). MGEs seem to largely explain the differences in susceptibility of closely related strains to different phages (Hussain et al., 2021). An important source of genetic variability is caused by temperate phages, which have been known for a long time to carry anti-phage functions. During lysogeny, phage survival is tied to host survival, providing a selective pressure for temperate phages to carry “moron” genes that are not essential for the phage but enhance the fitness of their host (Cumby et al., 2012). Prophage-encoded defense systems provide resistance to distant phages through diverse mechanisms, including modification of cell surface receptors (Uc-Mass et al., 2004), inhibition of DNA translocation (McGrath et al., 2002), premature transcription termination (Oberto et al., 1989), or abortive infection (Friedman et al., 2011; Montgomery et al., 2019; Owen et al., 2021; Snyder, 1995). Recent work highlighted how such prophage-encoded defense systems participate in inter-viral competition (Bondy-Denomy et al., 2016; Dedrick et al., 2017; Makarova et al., 2014).

Phages have evolved anti-defense strategies in response to defense systems. The resulting arms race is driving the
diversification and turnover of defense systems in bacteria and counter-defense systems in phages (Bernheim and Sorek, 2020). While bacteria can readily accumulate diverse defense systems in their genome within defense islands, the genome of bacteriophages is typically constrained by the size of the DNA that can be packaged into the capsid, limiting the number of systems they can carry. Genetic diversity in a specific phage clade is typically constrained to specific regions of the phage genome. In P2-like phages from Enterobacteria, two variable loci were described to include anti-phage genes (Nilsson et al., 2004; Odegrip et al., 2006). Anti-defense genes have also been found in variable regions of other phage genomes (Pavluk et al., 2014; Pinilla-Redondo et al., 2020).

Phage satellites represent another important class of MGEs. They hijack the capsid of helper phages to ensure their own propagation and frequently do so while inhibiting the propagation of their helper phage (Fillol-Salomon et al., 2020; O’Hara et al., 2017). As such, they are sometimes described as anti-phage systems, but whether satellites could provide defense against non-helper phages remains to be investigated.

Here, we show that hotspots of genetic diversity within the P2-like phage and P4-like satellite families constitute large reservoirs of anti-phage systems. These hotspots are small loci (~1–5 kb) with a high turnover of genetic material located between two conserved genes of the phage or satellite. This is in contrast to defense islands, which consist of chromosomal loci with a high concentration of defense systems and which can span tens of thousands of bases. We describe in more detail phage anti-restriction-induced system (PARIS), an abortive infection system that triggers growth arrest upon sensing an anti-restriction protein. We provide insights into the impact of these hotspots on inter-viral competition and describe diversity hotspots in prophages from distant bacterial species. These findings highlight how prophages constitute a reservoir of defense proteins and point to a strategy to uncover novel anti-phage systems.

## RESULTS

### P4-like phage satellites encode a hotspot for anti-phage systems

While investigating the determinants of gene essentiality in various *E. coli* strains, we previously identified a reverse-transcriptase (RT) associated with a SLATT-domain protein that is responsible for the essentiality of the exodeoxyribonuclease I SbcB in the *E. coli* strain H120 (Rousset et al., 2021). Since diverse bacterial RTs have recently been implicated in anti-phage defense (Bobonis et al., 2020a, 2020b; Gao et al., 2020; Mestre et al., 2020; Millman et al., 2020a), we investigated the genetic neighborhood of this system. Instead of being in a defense island, it was inserted in a P4-like phage satellite between the polarity suppression protein gene (psu) and the integrase (int). Since P4-like satellites are very prevalent in Enterobacteria, including 44% of *E. coli* isolates (de Sousa and Rocha, 2022), we inspected other *E. coli* strains and noticed that each P4-like element carried a different genetic system at the same locus (Figure 1A). These elements are adjacent to the cos packaging signal and found in either orientation. They include a variety of uncharacterized proteins as well as known defense proteins belonging to type-II and -III restriction-modification (RM) systems and retrons, two of which had previously been reported (Nouye et al., 1991; Kita et al., 1999). The well-characterized P4 satellite (NC_001609) carries at this position the non-essential genes gop, b, and cII, with gop and b likely acting as a toxin-antitoxin pair (Ghisotti et al., 1990). We identified 5,251 occurrences of this locus in >20,000 *E. coli* genomes (STAR Methods; Table S1A), together encoding >300 different gene arrangements. We analyzed 121 arrangements with at least five occurrences (together accounting for 94.4% of all loci) (Figure S1A) and identified 79 groups of genetic systems (Figure 1B; Table S1B). They can be broadly assigned to three categories: (1) known defense systems such as a type-III restriction enzyme, the type-II RM system EcoO109I (Kita et al., 1999), Kiwa (Doron et al., 2018), Septu (Doron et al., 2018), or SIR2 + HerA (Gao et al., 2020); (2) uncharacterized systems with protein domains that were previously associated with bacterial immunity such as SIR2 (Doron et al., 2018; Gao et al., 2020; Koyuncu et al., 2014), TIR (Ofir et al., 2021), and higher eukaryotes and prokaryotes nucleotide-binding (HEPN) (Anantharaman et al., 2013; Gao et al., 2020); and (3) systems encoding unannotated proteins or whose domain is currently not associated with anti-phage defense, such as haloacid dehydrogenase-like hydrolase (HAD), and various domains of unknown function (DUFs). Based on these observations, we hypothesized that the P4-encoded locus between *psu* and *int* may constitute a reservoir of antiviral systems that likely participate in inter-viral competition.

To test for anti-phage activity, we cloned the set of genes present in the canonical P4 satellite (NC_001609), as well as 18 systems encompassing the three categories above from strains of *Escherichia* and *Klebsiella* under the control of their native promoter on a low copy number vector (Tables S2A and S2B). We introduced them into *E. coli* K-12 MG1655 and challenged each resulting strain with an array of eight coliphages spanning most common phage families (key resources table). When compared with a control vector encoding a green fluorescent protein (GFP), seven systems provided robust and reproducible resistance to at least one phage (Figures 2 and S2; Table S2A).

We report that gop-β-cII forms an anti-phage system that protects against λ and P1. The gop gene product was shown to be toxic in the absence of β (Ghisotti et al., 1990), suggesting that the system may form a toxin-antitoxin system functioning through abortive infection. We also identified a RT from unknown group 5 (Sharifi and Ye, 2021; Toro and Nisa-Martínez, 2014) with a C-terminal nitrilase domain (Table S3) that is associated with a transmembrane effector. This system is reminiscent of a recently described anti-phage system called type-I DRT (Gao et al., 2020), although the RT belongs to a different clade. Validated systems also include a single-gene system with a TIR (toll/interleukin-1-receptor-like) domain protein. TIR domains are important determinants of eukaryotic immunity (Takeda and Akira, 2005) and evidence is emerging of a role for TIR domain proteins as bacterial immune factors as well: the Thoeris defense system carries a TIR protein which produces a second messenger from NAD+ to activate an associated cell-killing effector (Ofir et al., 2021), while some CBASS and Pycsar systems use a TIR domain to deplete cellular NAD+ to induce abortive infection (Morehouse et al., 2020; Tal et al., 2021b). Here, the...
TIR protein also harbors C-terminal tetratricopeptide repeats (TPRs) and a central STAND NTPase domain (Leipe et al., 2004) that shares homology with eukaryotic NOD-like receptors (NLRs). In Eukaryotes, the STAND domain is involved in programmed cell death and innate immunity by providing ATP/GTP-dependent oligomerization upon signal sensing, leading to downstream signaling (Kufer et al., 2005). In bacteria, STAND NTPases were recently described in the AVAST family of defense systems. See also Figures S1 and S3 and Tables S1 and S2.
systems (Gao et al., 2020). Here, TPRs may sense phage infection, inducing ATP- or GTP-dependent oligomerization and activation of the TIR effector to degrade cellular NAD⁺. Other proteins from the hotspot also harbor a STAND NTPase domain coupled to C-terminal repeats, but where the N-terminal TIR effector domain can be replaced by a sirtuin (SIR2) or nuclease (Mrr_cat or BpuSI_N) domain (Table S1B). Taken together, our findings expand the scope of this family of anti-phage systems, which shares striking homology with NLRs from plants and animals.

We also describe two occurrences of an ATPase associated with a DUF4435 protein (described below as PARIS) and two systems that specifically inhibit the growth of phage T7: an ATP-dependent helicase associated with a DUF2290 protein and a HAD-like hydrolase associated with a transmembrane protein. In currently known anti-phage systems, transmembrane proteins are believed to be effectors triggering cell suicide upon infection (Duncan-Lowey et al., 2021; Millman et al., 2020b; Snyder, 1995; Tal et al., 2021b), suggesting that this system could work through abortive infection.

Taken together, our findings show that P4-like satellites carry a hotspot that constitutes a reservoir of anti-phage defense systems. While our bioinformatic search is limited to E. coli genomes, P4-like elements identified in other Enterobacteriaceae also carry known and candidate defense genes at the same locus (Figure S3).

PARIS triggers growth arrest upon sensing an anti-restriction protein

We further investigated a system comprising an AAA+ ATPase associated with a DUF4435 protein (Figure 2) that we renamed ariA and ariB, respectively (see below). We detected ariAB in 5.2% of bacterial and archaeal genomes from diverse clades (Figures S4A and S4B; Tesson et al., 2021), either as a two-gene cassette as described here or as a single-gene fusion comprising both domains, indicating strong evidence of tight functional interaction (Enright et al., 1999). Sequence analysis of AriB using HHpred identified a ~60-amino acid segment with a weak but significant match (E value = 0.02) to the topoisomerase-primase (TOPRIM) domain of OLD family nucleases (Schiltz et al., 2019, 2020). Interestingly, the association of an ATPase with a TOPRIM domain has previously been found in other defense systems such as P2 old that protects against λ (Myung and Calendar, 1995) and GajA from the Gabija system (Cheng et al., 2021; Doron et al., 2018), suggesting that these proteins form a large family of defense proteins. While the system from E. coli B185 provided robust defense when expressed from its natural promoter, a distant homolog from E. coli O42 initially showed no activity against our phage panel. To ensure that this absence of phenotype was not due to a lack of expression, we cloned this system under the control of an anhydrotetracycline (aTc)-inducible promoter. Upon induction, anti-phage activity was observed for a large range of phages and with a different but overlapping defense profile compared with the system found in E. coli B185 (Figure 2).

We then focused on the system from E. coli B185 for subsequent experiments. Deletion of either ariA or ariB was non-toxic and abolished defense, excluding the hypothesis of a toxin-antitoxin system and showing that both components are required for activity (Figures S4C–S4E). Mutations of the predicted ATPase and TOPRIM catalytic residues also abolished defense. We monitored the growth of cells carrying the system or a control plasmid during infection by phage T7 at a low (0.005) or high (5) multiplicity of infection (MOI). In the presence of a control plasmid, infection led to population collapse regardless of the MOI (Figure 3A). In contrast, the system provided partial resistance to T7 at a low MOI but led to a growth halt after infection at a high MOI. In addition, the number of infected cells that released viable phage was reduced by ~15-fold (95% CI: [12.1;18.5]) in the presence of the system, as estimated by a center of infection assay (STAR Methods; Table S4A). Therefore, ariAB seem to halt growth in infected cells, preventing the

---

**Figure 2.** P4-encoded hotspots encode a variety of anti-phage systems

Phage resistance heatmap of the validated defense systems shows the mean fold resistance of three independent replicates against a panel of eight phages (key resources table). Genes are colored by protein family. Genome accession numbers are provided in Table S2A. Systems are under the control of their native promoters, with the exception of PARIS-1, which was only active when expressed from a ptet promoter (pFD237) in the presence of anhydrotetracycline (aTc). Defense was measured at 37°C, with the exception of the RT-nitrilase + 1TM system, which was measured at room temperature. RT, reverse-transcriptase; TIR, Toll/interleukin-1 receptor; HAD, haloacid dehydrogenase-like; DUF, domain of unknown function; TM, transmembrane helix. See also key resources table, Figure S2, and Tables S2A and S3.
completion of the phage cycle, a typical feature of abortive infection systems (Lopatina et al., 2020).

We aimed at identifying the viral trigger by isolating phages that were able to overcome this system, reasoning that mutations in the trigger component would enable phages to escape defense. Whole genome sequencing of four T7 mutants with a restored infectivity (Figure 3B) revealed that all of them carried a mutation (F54V) in gene 0.3, which encodes Ocr, an anti-restriction protein that inhibits RM and BREX systems by mimicking the structure of DNA (Isaev et al., 2020; Studier, 1975). In the presence of ariAB, the transformation efficiency of an Ocr-expressing plasmid was ~100-fold lower than in the absence of the system and yielded only small and sick colonies, while there was no difference when transforming a GFP-encoding plasmid as a control (Figure 3C). This confirmed the toxicity of Ocr in the presence of this system. We can therefore describe it as an anti-anti-restriction system, a remarkable evolutionary strategy where cells have evolved to use a counter-defense.
protein as a trigger for abortive infection. This strategy allows cells to undergo self-sacrifice only when an incoming phage can overcome the first line of defense (e.g., a RM or BREX system), therefore maximizing both cellular and population-level survival. As such, we renamed this system PARIS and its components arIA and arIB (anti-restriction-induced A and B).

During infection by T7, the Ocr protein inhibits EcoKI, the RM system that is naturally present in E. coli K-12 MG1655, preventing both restriction and methylation of T7 DNA (Studier, 1975). Therefore, T7 mutants isolated on E. coli K-12 MG1655 expressing PARIS must be able to overcome both PARIS and EcoKI. We then wondered if the OcrF54V mutant that escaped PARIS was still able to block EcoKI. To investigate this, a plasmid carrying an EcoKI restriction motif was extracted either from the strain DH10B (that lacks EcoKI) to yield unmethylated DNA or from strain MG1655 to yield methylated DNA. Transformation efficiency was then measured in MG1655 cells expressing Ocr, OcrF54V, or a control plasmid. While unmethylated plasmid was clearly restricted by EcoKI, both unmethylated and methylated plasmid DNA could be transformed with similar efficiency when recipient cells expressed Ocr or OcrF54V (Figure 3D). This shows that OcrF54V can indeed still block EcoKI.

We then reasoned that the evolutionary pressure on Ocr to maintain EcoKI inhibition should be absent in EcoKI- cells. To investigate this, we introduced PARIS in DH10B cells (that lack EcoKI) and selected another set of T7 mutants that escape PARIS. Sequencing the ocr gene indeed revealed inactivating frameshifts (ΔT35, ΔG209) or non-sense mutations (247C > T or 265C > T) in 15/16 of isolated mutants (Table S4B). As expected, an ocr(ΔT35) mutant of T7 had a reduced efficiency of plaquing on MG1655, but blocking EcoKI expression with dCas9 restored infectivity (Figure 3E). Altogether, our results show that PARIS acts as a second line of defense when a phage is able to inactivate RM systems (Figure 3F), but that anti-restriction proteins can evolve to bypass PARIS while maintaining RM inhibition; e.g., with the F54V mutation in Ocr. A sequence search against the NCBI Viral Database (Hatcher et al., 2017) revealed that the F54 residue is conserved in 93% of Ocr homologs, suggesting that the F54V mutation might have disadvantages in some conditions. Accordingly, a F54A mutation was previously shown to abrogate Ocr dimerization that is required for full inhibition of type-I RM and BREX (Isaev et al., 2020; Zavilgelsky and Kotova, 2014).

A hotspot in P2-like phages encodes diverse anti-phage systems

P4 is a phage satellite that lacks structural genes encoding capsid and tail proteins. Instead, it hijacks capsids from helper phages of the P2 family and uses the P2 terminase to package its own genome into modified P2 capsids (Lindqvist et al., 1993). As a consequence, P4 and P2 share a core cos packaging signal (Ziermann and Calendar, 1990). The fact that the defense hotspot is directly adjacent to the cos site in P4-like phages prompted us to inspect the cos-proximal region in P2-like phages, which is located between the replication protein gpA and the portal protein gpQ (Figure 4A). The P2 reference genome (NC_001895) encodes the accessory phage exclusion genes old and tin at this locus (Christie and Calendar, 2016). We performed a systematic analysis of this locus in all E. coli genomes which showed that this region is another hotspot for genetic diversity as previously suggested (Odegrip et al., 2006), with 1,650 different gene arrangements detected from 18,150 occurrences of this locus (Figure 4B; Table S5A). We curated the arrangements occurring at least 10 times (together accounting for 82% of all loci) (Figure S1B) and identified 169 genetic systems (Table S5B) comprising genes encoding known defense proteins such as retrons and type-III restriction endonucleases, as well as genes likely involved in other functions. For instance, around a fourth of P2-like phages encode a putative plasmid partitioning system that perhaps allows these prophages to be maintained as plasmids, while around 2% encode a cytolysin distending toxin which likely participates in bacterial virulence (Johnson and Lior, 1988). A few systems are common between P4- and P2-encoded hotspots but at different frequencies, while most systems are found in one hotspot but not in the other. Many P2-like hotspots contain more than one system and they tend to be larger than those of the P4-like (3.2 versus 2.6 kb on average) (Figure S1C), likely reflecting the pressure faced by the latter to package their genome into a capsid of reduced size which can only accommodate ~12 kb (Shore et al., 1978). Accordingly, both phage families encode short systems and lack larger systems such as CRISPR-Cas, BREX, and type-I RM that can be found in bacterial genomes or plasmids.

We cloned eight additional systems from the P2-encoded hotspot and tested their activity against the eight phages described above. In addition, we also cloned the old and tin genes from bacteriophage P2 to confirm their previously reported anti-phage activity (Christie and Calendar, 2016). The old-tin genes provided immunity against a broad range of phages, and we observed defensive activity for three other systems (Figure 5A). The first system consists of a single protein with a DUF4238 domain (whose molecular function is currently unknown) that provides strong resistance against T7. The second system comprises three genes that have no clear predicted domain and protect against P1. Finally, the last system comprises a DNA-processing chain A protein (DprA) associated with a phosphoribosyltransferase (PRTase) and protects against phages T7 and λ. DprA-like proteins are involved in DNA transformation in naturally competent bacteria by binding to single-stranded DNA and interacting with RecA (Mortier-Barrière et al., 2007). Non-competent species also carry DprA-like proteins whose role has remained elusive. Our finding now provides another function to DprA-like proteins in non-competent bacteria. Some PRTase proteins act as probable effectors in bacterial retrons although their mode of action is unknown (Mestre et al., 2020; Millman et al., 2020a). This suggests that DprA could sense phage DNA and activate the PRTase effector, perhaps to trigger cell suicide in an abortive infection mechanism. This hypothesis is supported by the role of DprA homologs as antitoxins in the shosTA, syrTA, and rqHI toxin-antitoxin systems (Kimeiman et al., 2012; Russell and Mulvey, 2015; Sberro et al., 2013). Taken together, our results suggest that both P4 satellites and P2-like phages carry a hotspot of genetic diversity adjacent to the cos site, with a large variety of genes involved in anti-phage defense.

We further investigated the ability of P2-like phages to confer protection to their host. We isolated four P2-like phages in E. coli C, a restriction-less strain traditionally used in bacteriophage isolation (key resources table; Figure S5; Bertani and Weigle, 1953). We then selected E. coli C lysogens and tested their
resistance against our phage panel. The P2-like phage AC1 protected its host against λ and LF82_P8 (Figure 5B). At the defense hotspot, AC1 carries a TIR protein that may generate a nucleotide messenger to activate the associated transmembrane proteins (Figure 5C; Ofir et al., 2021). We cloned the three genes present at the AC1 locus on a plasmid that we introduced in E. coli and confirmed their activity against λ and LF82_P8 (Figure 5D).

When tested against P2-like phages, this system also provided resistance against P2 itself (Figure 5E), showing that diversity hotspots can participate in inter-viral competition, not only between distant phages but also at a short scale between closely related P2-like phages. Altogether, this confirms that genetic systems carried at the hotspot can provide resistance against diverse phages in their natural context.

Role of defense systems in P2-P4 interactions

The presence of defense systems in P4-like satellites can provide a competitive advantage to a host by providing protection...
against phages. The interests of the bacterium and the phage satellite are however not perfectly aligned. Indeed, when the cell is infected by a phage from the P2 family, P4-like satellites can ensure their own propagation by packaging their DNA into modified P2-like capsids, while the cell dies. We can thus hypothesize that defense systems found in P4-like satellites will usually not restrict the P2-like phages they can hijack or do so while ensuring the propagation of the P4 element. In fact, the only phage in our panel against which no defense could be detected was the P2-like phage 186cIts (Figure 2). All the P4-encoded systems described above (Tables S2A and S2B) were introduced into E. coli C, and their defense activity was tested against P2 and the four newly isolated P2-like phages. Out of the 18 tested systems, only the TIR-NLR system provided protection (Figure 6A). This is in contrast with the seven systems described in Figure 2 that provided defense against the phages of our panel and the fact that non-P2-like phages were all restricted by several systems. These results suggest that defense systems carried by P4-like satellites are generally permissive to P2-like phages.

We further investigated whether the defense action of the TIR-NLR system against P2-like phages could occur in a timeframe that still enables P4 packaging and transduction. E. coli C cells carrying the P420 plasmid, a P4 derivative with a kanamycin resistance gene (Kahn and Helinski, 1978), were infected by P2 in the presence or absence of the TIR-NLR system and P420 titers were measured in the lysate. While P420 particles were still produced in the presence of the defense system, TIR-NLR reduced P420 titers to a similar extent by which it blocked P2 (Figure 6B). This shows that while the TIR-NLR system protects against a broad range of virulent phages, this comes at the cost of limiting the transduction of P4 by P2. Nevertheless, the natural helper phages of TIR-NLR-expressing satellites may not be affected by the system, as is the case for the P2-like phage 186cIts (Figure 2).

P4-like satellites can be beneficial to their bacterial host by protecting them from virulent phages. We further hypothesized that by blocking competing virulent phages, a satellite might be beneficial to its helper phage as well. To test this hypothesis, we co-infected E. coli C cells carrying P4 with a 1:1 mixture of phages P2 and LF82_P8 in the presence or absence of the
PARIS system on a plasmid (STAR Methods). In the absence of PARIS, LF82_P8 hindered the propagation of P2 (Figure 6C). In contrast, in the presence of PARIS, LF82_P8 was inhibited and P2 titers were close to those obtained during infection by P2 alone. In addition, PARIS also protected cells from the virulent phage (Figure 6D), thereby favoring lysogeny by P2. We observed \(10^9\) more colony-forming units after co-infection in the presence of PARIS than in its absence, among which \(\sim 15\%\) were P2 lysogens (Figure S6). These results show how P4 satellites, which are traditionally seen as parasites of P2, might in fact be beneficial to their helper phages by the selective action of their antiviral systems.

**Prophage-encoded antiviral hotspots in other species**

Finally, we wondered whether defense hotspots also exist in prophages belonging to more distant bacterial species. We were able to identify at least two more occurrences of such hotspots through manual examination of loci that carry some of the systems described here. The first hotspot is prevalent in Vibrionaceae, between the cl repressor and the integrase of a P2-related phage that is similar to *Vibrio cholerae* phage K139 (Kapfhammer et al., 2002) and *Aeromonas* phage ΦO18P (Figure 7A; Beilstein and Dreiseikelmann, 2008). The second one is encoded on a prophage from *Bacillus* between a transcriptional regulator and a peptidase adjacent to the integrase (Figure 7B). In both cases, hotspots encode known defense systems such as RM, Septu as well as the Helicase + DUF2290 system, and PARIS described above (Table S6). While most proteins in these loci have currently not been linked to bacterial immunity, many of them have known defense domains, such as nucleases, kinases, and ATPases. This analysis reveals that prophage-encoded diversity hotspots are present in various bacterial organisms and likely constitute a significant reservoir of new antiviral systems across diverse phyla.

**DISCUSSION**

Our findings showcase a stunning diversity of genetic systems in phages and satellites, with more than 300 unique gene arrangements detected in the P4 hotspot alone. This diversity is matched by the presence of highly diverse defense systems in P2-like phages in the only homologous region between the two types of elements, the cos site. Both types of elements include...
variants of known systems, or systems carrying combinations of domains previously associated to defense phenotypes, as well as systems with unknown domains or that were not previously associated to defense phenotypes. More than a third of the systems that we cloned showed anti-phage activity against our phages, T4, with the two PARIS systems likely reflecting the changing selective pressure to be unveiled. We found that PARIS also protects against other phages than T7, such as T4, with the two PARIS systems that we assayed offering varying levels of protection (Figure 2). Although the trigger of PARIS remains unknown for those phages, T4 is known to carry the anti-restriction protein Arn, which encodes the RecBCD inhibitor Gam, both of which are likely mirrors of the diversity of phages encountered by E. coli and the anti-defense strategies they deploy.

It is interesting to consider the molecular mechanisms driving the dramatic rate of gene exchange, specifically at these loci. In both P2-like phages and P4-like satellites, the hotspots are directly adjacent to the cos sequence. It is tempting to speculate that cohesive ends generated after cleavage by the terminase or made available immediately after injection of the phage DNA into the cytoplasm could provide a substrate for recombination upon co-infection by several P2-like phages or P4-like satellites, or upon co-induction of resident prophages. A second recombination event mediated by homologous recombination would result upon co-induction of resident prophages. A second recombination event mediated by homologous recombination would result upon co-induction of resident prophages. A second recombination event mediated by homologous recombination would result upon co-induction of resident prophages.

Among the many systems, we describe in more detail PARIS, an abortive infection system that senses the T7 Ocr anti-restriction protein (Figure 3). PARIS seems conceptually similar to the anticodon nuclease PrrC, which acts as a second line of defense by sensing the inhibition of the EcoKII RM system by Stp, a peptide encoded by phage T4 (Penner et al., 1995). Both systems represent a remarkable evolutionary strategy where cells commit suicide only when a first line of defense is compromised, a strategy also recently described in an antiviral retron that senses RecBCD inhibition by a phage protein (Millman et al., 2020a). However, PARIS is distinct in the sense that it does not require the presence of EcoKII or any other type-I RM or BREX system in the cell, as shown by its defensive activity in DH10B cells (Figure 3E). We can thus rule out that PARIS guards EcoKII, but the exact mechanism of anti-restriction sensing by PARIS remains to be unveiled. We found that PARIS also protects against other phages than T7, such as λ and T4, with the two PARIS systems that we assayed offering varying levels of protection (Figure 2). Although the trigger of PARIS remains unknown for those phages, T4 is known to carry the anti-restriction protein Arn, while λ encodes the RecBCD inhibitor Gam, both of which are DNA mimics like Ocr (Court et al., 2007; Ho et al., 2014). Future studies should address whether PARIS is able to sense phage-encoded DNA mimics in general.

The differences in defense profiles provided by different anti-phage systems likely reflects the changing selective pressure.

**Figure 7.** Hotspots for anti-phage systems encoded on other prophage genomes

(A and B) Genomic view of hotspots encoded on prophages from Vibrionales (A) and Bacilli (B). Phage genes are shown with different shades of blue. Gray shades show the percentage of identity between homologous proteins from different genomes. Genome accession numbers and positions are shown on the left. See also Table S6.

---

[Note: The image contains diagrams and metadata that are not transcribed here due to the nature of the task.]
imposed by predation by diverse bacteriophages against which no universal defense exists. Our results point toward a role of P2- and P4-encoded systems in the protection against phages of other families. Defense systems found at the P2 hotspot did not protect against P2-like phages, with the exception of a system from phage AC1, which provided protection against P2 (Figure 5).

Phage satellites were previously described to frequently inhibit their helper phage (Filiol-Salom et al., 2020; O’Hara et al., 2017), but the discovery of defense systems active against a broad range of phages on P4-like satellites raises the question of how they might affect the relationship between the phage, the satellite, and their common host. The existence of abortive infection systems in P4-like elements suggests that it is best for them to kill their host rather than allow an epidemic of phages they cannot highjack. Interestingly, the benefit provided to the host might extend to the helper phage. This is because the selective defense provided by a P4 element favors the propagation of helper P2 phages, either as lysogens or virions, during co-infection with a virulent phage (Figure 6). This challenges the paradigm of the satellite as a phage parasite by showing that in some circumstances the interactions might be mutualistic. The P2 phage propagates the P4 element, while the P4 element gives an advantage to P2 during competition with virulent phages. The diversity of defense systems provided by P4-like elements can then be seen as a rich genetic resource that P2 can harness to protect itself, and its host, against virulent phages without the need to encode all these elements in its own, restricted size, genome.

The integration of temperate phages, their satellites (this work), and ICEs (Johnson et al., 2020; LeGault et al., 2021) carrying defense systems provides a mechanism for the genesis and evolution of defense islands in bacterial genomes. These mobile elements tend to integrate at a small number of chromosomal hotspots (Oliveira et al., 2017). Their inactivation by mutation results in rapid gene loss until only the functions adaptive to the bacterial host remain (Touchon et al., 2014). In fact, we have found occurrences of P4-like satellites that are integrated near defense islands or ICEs that carry defense genes (Figure S7). The high turnover of mobile elements in a few chromosomal hotspots can lead to the accumulation of defense systems in specific loci. Accordingly, recent studies showed how the majority of the flexible genome among close relatives of Vibrio consists of antiphage defense elements (Hussain et al., 2021; Piel et al., 2021). These regions may also be hotbeds for recombination and mutation processes, resulting in novel combinations of protein motifs that may in some cases result in novel defense systems and favoring the genetic entanglement between defense systems, viruses, and MGEs in general (Koonin et al., 2020; Rocha and Bikard, 2022). The bacterial chromosome is much less constrained than those of phages and can thus accumulate many more systems in the same hotspot. In this hypothesis, bacterial defense islands against mobile elements result from the accumulation of defense systems from mobile elements themselves.

Finally, the identification of defense hotspots adjacent to conserved anchor genes provides an alternative strategy to uncover novel defense systems. Unlike previous systematic approaches that rely on the proximity of systems in relatively large loci (Doron et al., 2018; Gao et al., 2020), defense hotspots should be identifiable in metagenomic and virome data with small contigs. We therefore anticipate that this strategy will help uncover more antiviral systems involved in the conflicts between MGEs.

Limitations of the study
We detected the presence of defense hotspots in species other than E. coli, but a systematic analysis of the prevalence of such hotspots in phages and satellites across the bacterial kingdom will require a dedicated bioinformatics study that is beyond the scope of this work. In addition, the fraction of defensive systems encoded in the hotspots that we report is likely underestimated due to the small size of our phage panel.

Our findings argue for a role of antiviral hotspots carried by satellites, but more work will be required to ascertain their impact on the eco-evolutionary dynamics of host-phage-satellite interactions in natural E. coli hosts. We observed a trend suggesting that defense systems carried by P4 satellites generally do not restrict P2-like phages, but this conclusion is only based on a handful of interactions. The study of more systems and phages will be necessary to draw stronger conclusions. It also remains unclear under what circumstances phage satellites will be selected to block or not block the propagation of their helper phage. More work will also be necessary to establish the fitness benefit that P2 phages can obtain from mobilizing P4 elements and the anti-phage systems they carry. Transduction of P4-like elements by P2-like phages spreads systems that will favor subsequent infection by the latter, but it remains unclear if and when these benefits outweigh the costs paid by the phage in terms of its effective burst size.

Finally, future studies will ascertain how anti-phage systems are captured in defense islands, the role that defense hotspots might play in this process, and the extent of genetic exchanges between phages, satellites, and the host.

STAR Methods
Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bacteria
  - Phages
- METHOD DETAILS
  - Cloning candidate defense systems
  - Phage plaque assays
  - Time course infection experiments
  - Efficiency of centers of infection
  - Isolation and sequencing of mutant phages
  - Transformation assays
  - CRISPRi-mediated EcoKl knockdown
  - Effect of TIR-NLR on P2-P4 interaction
  - P4-Kan construction
  - Competition between P2 and LF82_P8 in the presence of PARIS
Supplemental information can be found online at https://doi.org/10.1016/j.chom.2022.02.018.

ACKNOWLEDGMENTS

We thank Erick Demarain for providing E. coli strains and Sylvain Briaise for providing K. pneumoniae strains. We also thank Laurent Debarbieux and Mathieu de Jode for providing phages CLB_P2, LF82_P8, and AL505_P2, the P2M platform (Institut Pasteur, Paris, France) for genome sequencing, and Ali-thieu de Jode for providing phages CLB_P2, LF82_P8, and AL505_P2, the providing chom.2022.02.018.

SUPPLEMENTAL INFORMATION

We also thank Laurent Debarbieux and Ma-l Laurenceau for their contribution to the PARIS acronym. This work was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement no. 677823), by the French Government’s Investisse- ment d’Avenir program, and by Laboratoire d’Excellence “Integrative Biology of Emerging Infectious Diseases” (ANR-10-LABX-62-IBEID). F.R. is supported by a doctoral scholarship from École normale supérieur. E.P.C.R. was supported by the results and wrote the manuscript.

AUTHOR CONTRIBUTIONS


DECLARATION OF INTERESTS

D.B. is a founder of Eligo Bioscience and a member of its scientific advisory board.

REFERENCES


## STAR★METHODS

### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial and virus strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em> K-12 MG1655</td>
<td>Gift of Mazel lab</td>
<td>RefSeq: NC_000913</td>
</tr>
<tr>
<td><em>E. coli</em> C</td>
<td>Collection de l'Institut Pasteur</td>
<td>CIP 104337</td>
</tr>
<tr>
<td><em>E. coli</em> DH10B</td>
<td>Invitrogen</td>
<td>EC0113</td>
</tr>
<tr>
<td><em>E. coli</em> ACE1</td>
<td>Calvo-Villamanán et al., 2020</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Escherichia phage</em> λ</td>
<td>Gift of Luciano Marraffini</td>
<td>RefSeq: NC_001416</td>
</tr>
<tr>
<td><em>Escherichia phage</em> T4</td>
<td>Gift of Laurent Debarbieux</td>
<td>RefSeq: NC_000866</td>
</tr>
<tr>
<td><em>Escherichia phage</em> P1</td>
<td>Gift of Jean-Marc Ghigo</td>
<td>RefSeq: NC_005856</td>
</tr>
<tr>
<td><em>Escherichia phage</em> 186cIts</td>
<td>Gift of Keith Shearwin</td>
<td>RefSeq: NC_001317</td>
</tr>
<tr>
<td><em>Escherichia phage</em> CLB_P2</td>
<td>Maura et al., 2012</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Escherichia phage</em> LF82_P8</td>
<td>Galtier et al., 2017</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Escherichia phage</em> AL505_P2</td>
<td>Galtier et al., 2016</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Escherichia phage</em> T7</td>
<td>Félix d'Hérelle Reference Center for Bacterial Viruses</td>
<td>RefSeq: NC_001604</td>
</tr>
<tr>
<td><em>Escherichia coli</em> bacteriophage P4 sid1</td>
<td>ATCC</td>
<td>ATCC29746-B1</td>
</tr>
<tr>
<td><em>Escherichia phage</em> P2</td>
<td>ATCC</td>
<td>ATCC-29746</td>
</tr>
<tr>
<td><em>Escherichia phage</em> P2_AC1</td>
<td>This study</td>
<td>ENA: SAMEA9990737</td>
</tr>
<tr>
<td><em>Escherichia phage</em> P2_DC1</td>
<td>This study</td>
<td>ENA: SAMEA9990738</td>
</tr>
<tr>
<td><em>Escherichia phage</em> P2_SIDE7</td>
<td>This study</td>
<td>ENA: SAMEA9990739</td>
</tr>
<tr>
<td><em>Escherichia phage</em> P2_SIAC10</td>
<td>This study</td>
<td>ENA: SAMEA9990740</td>
</tr>
<tr>
<td><strong>Chemicals, peptides, and recombinant proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kanamycin</td>
<td>Sigma</td>
<td>Cat # K0254</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>Euromedex</td>
<td>Cat # 3886-C</td>
</tr>
<tr>
<td>Carbencillin</td>
<td>Euromedex</td>
<td>Cat # 1039-A</td>
</tr>
<tr>
<td>Anhydrotetracycline hydrochloride</td>
<td>ThermoFisher</td>
<td>Cat # 233131000</td>
</tr>
<tr>
<td>Phusion Polymerase</td>
<td>ThermoFisher</td>
<td>Cat # F530L</td>
</tr>
<tr>
<td>TURBO DNase</td>
<td>ThermoFisher</td>
<td>Cat # AM2238</td>
</tr>
<tr>
<td>Proteinase K</td>
<td>Eurobio</td>
<td>Cat # GEXPRK01B5</td>
</tr>
<tr>
<td>Phenol-chloroform-isooamylalcohol solution</td>
<td>Sigma</td>
<td>Cat # P3803</td>
</tr>
<tr>
<td>Chloroform</td>
<td>Sigma</td>
<td>Cat # C2432</td>
</tr>
<tr>
<td><strong>Critical commercial assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NucleoSpin Plasmid</td>
<td>Macherey-Nagel</td>
<td>Cat # 740588.50</td>
</tr>
<tr>
<td><strong>Oligonucleotides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All the DNA oligonucleotides are listed in Tables S2A, S2B, and S7B and method details</td>
<td>Eurofins Genomics</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Recombinant DNA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasmid psgRNAc</td>
<td>Cui et al., 2018</td>
<td>Addgene Cat #114006</td>
</tr>
<tr>
<td>All the plasmids are listed and described in Tables S2A, S2B, and S7B and method details</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Software and algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Python 3.8.5</td>
<td>Python Software Foundation</td>
<td><a href="https://www.python.org/downloads/release/python-38/">https://www.python.org/downloads/release/python-38/</a></td>
</tr>
<tr>
<td>MMseqs2</td>
<td>Steinegger and Söding, 2017</td>
<td><a href="https://github.com/soedinglab/MMseqs2">https://github.com/soedinglab/MMseqs2</a></td>
</tr>
<tr>
<td>clinker &amp; clustermap.js</td>
<td>Gilchrist and Chooi, 2021</td>
<td><a href="https://github.com/gamcil/clinker">https://github.com/gamcil/clinker</a></td>
</tr>
<tr>
<td>HHpred</td>
<td>Zimmermann et al., 2018</td>
<td><a href="https://toolkit.tuebingen.mpg.de/tools/hhpred">https://toolkit.tuebingen.mpg.de/tools/hhpred</a></td>
</tr>
</tbody>
</table>

(Continued on next page)
RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, David Bikard (david.bikard@pasteur.fr).

Materials availability
All materials generated for this study are available upon request and without restrictions from the lead contact, David Bikard. All data is available in the main text or the supplemental information.

Data and code availability
- *E. coli* genome sequences used in this study are freely available from RefSeq.
- This paper does not report original code. All programs used to analyze genomes were previously reported and are freely available online (see key resources table).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacteria
*E. coli* K-12 MG1655, *E. coli* C or *E. coli* DH10B were grown at 37°C in Lysogeny Broth (LB) medium, unless stated otherwise. Kanamycin (Kan, Sigma), carbenicillin (Carb, Euromedex) and chloramphenicol (Cm, Euromedex) were used at 50 μg/mL, 100 μg/mL and 20 μg/mL respectively.

Phages
Phages were amplified on *E. coli* K-12 MG1655 (λ, T4, P1, 186Cltts, CLB_P2, LF82_P8, AL505_P2, T7) or *E. coli* C (P2, AC1, DC1, SIDE7, SIAC10; see also key resources table). Phage stocks were amplified by mixing 100 μL of an overnight culture of *E. coli* with 10 μL of phage stock solution (either pure or diluted) and 5 mL of warm (~50°C) LB + CaCl₂ 5 mM + 0.5% agar and poured on LB + CaCl₂ 5 mM + 1% agar plates. Plates on which confluent lysis was observed were used to recover the top-agar layer in 1mL of PBS and transfer it to a 50 mL conical tube. The top agar was then disrupted by vortexing until broken into small pieces. Tubes were then left to incubate 10 min at room temperature before centrifugation at 3,000 g for 5 min. Finally, the supernatant was recovered.

P2-like phages AC1, DC1, SIDE7 and SIAC10 were isolated as follows. *E. coli* from Eligo Bioscience’s strain collection were grown overnight in deep-well 96-well plates in 1 mL LB medium with agitation and at 37°C. The next day, cultures were diluted 1:100 into 1 mL LB medium and allowed to re-grow at 37°C with agitation for 3 h. The cultures were diluted 1:10 into 1 mL LB medium with or without mitomycin C (Sigma M4287-2MG final concentration of 0.1 μg/mL). After 6 h of incubation with agitation at 37°C, cultures were filtered with a 0.45 μm filter and the lysates were serially diluted and spotted onto lawns of *E. coli* C. Plaques appeared after overnight incubation at 37°C and were reisolated for a total of three times. AC1 and DC1 were obtained from cultures induced with mitomycin while SIDE7 and SIAC10 were obtained from the supernatant of uninduced cultures. Phage DNA was extracted and sequenced as described below. Genomes were assembled using SPAdes 3.15.3 (Bankevich et al., 2012) and deposited with the following accession numbers: ENA: SAMEA9990737 (AC1), ENA: SAMEA9990738 (DC1), ENA: SAMEA9990740 (SIAC10) and ENA: SAMEA9990739 (SIDE7).

METHOD DETAILS

Cloning candidate defense systems
Systems were amplified from the source strains indicated in Tables S2A and S2B, with the exception of 3 systems which were synthesized (Twist Bioscience): the active system from *E. coli* 2862600 as well as two systems from *E. coli* HVH3 and O157:H7 FRIK944.
for which we did not detect any activity against our phage panel (Table S2B). All systems were amplified with their native promoters by Phusion PCR (Thermo Fisher) using primers listed in Tables S2A and S2B. We used as vector pFR66, a low-copy plasmid with a pSC101 origin of replication, a kanamycin resistance cassette and a superfolder GFP gene (DNA sequence is provided in Table S7A). pFR66 was made by linear PCR with primers 5'-TTTTGCTCCTACAATGTC-3' and 5'-CCAGGCATCAAATTAAAAC GAAAGGCTCAGTCAG-3'. The GFP gene was replaced by each candidate system using the Gibson method (Gibson et al., 2009). All systems were transformed into electrocompetent E. coli K-12 MG1655 cells which were prepared as follows: following 100-fold dilution of an overnight culture in 200 mL of LB, cells were grown to OD ~1, harvested (4,000 g – 7 min) and washed three times in ice cold water, before concentration in ~300 μL of 10% glycerol. One microliter of dialyzed Gibson assembly product was transformed into 20 μL of cells. After 1 h of recovery at 37°C in LB medium, transformants were selected on LB + Kan plates. Mutants of PARIS-2 were constructed using each candidate system using the Gibson method (Gibson et al., 2009) after PCR amplification using primers listed in Table S7B. The TIR-NLR and PARIS-2 systems were also cloned on a derivative of pFR66 providing resistance against chloramphenicol (pFD232), giving pFD233 (TIR-NLR) and pFD235 (PARIS-2) respectively. The PARIS-1 system was cloned under the control of a T7 inducible Ptet promoter on a derivative of pFR66 leading to pFD237. All constructions were verified using Sanger sequencing.

**Phage plaque assays**

_E. coli_ K-12 MG1655 or _E. coli_ C strains carrying each of the systems or the control plasmid pFR66 were grown overnight in LB + Kan. Bacterial lawns were prepared by mixing 250 μL of a stationary culture with 62.5 μL of CaCl₂ 1 M and 12.5 μL of LB + 0.5% agar and the mixture was poured onto large square plates (12 x 12 cm) of LB + Kan. Serial dilutions of high-titer (>10⁸ pfu/mL) stocks of phages λ, T4, P1, 186C1ts (a thermosensitive variant of 186), CLB_P2 (Maura et al., 2012), LF82_P8 (Galtier et al., 2017) and AL505_P2 (Galtier et al., 2016) were spotted on each plate and incubated overnight at 37°C. For phage T7, plates were incubated overnight at room temperature. For the RT-nitrilase + 1TM system, all incubations were performed at room temperature. Information related to our phage panel is provided in key resources table. The next day, plaques were counted and the fold resistance was measured as the number of plaques in the control plate divided by the number of plaques in the presence of each system. When plaques were too small to be counted individually, we considered the most concentrated dilution where no plaque was visible as having a single plaque. The complete list of validated systems is provided in Table S2A, while additional tested systems are provided in Table S2B.

**Time course infection experiments**

Overnight cultures of K-12 MG1655 cells carrying a control plasmid or a plasmid encoding PARIS from _E. coli_ B185 were diluted to OD ~ 0.04 in LB + Kan and arrayed in a 96-well plate. Growth was then monitored in three replicates every 5 min on an Infinite M200Pro (Tecan) at 37°C with shaking. When OD reached ~0.2, cells were either kept uninfected or infected with ~2.10⁸ pfu (MOI ~ 5) or ~2.10⁹ pfus (MOI ~ 0.005) of phage T7. Growth was then monitored for 6 h post-infection.

**Efficiency of centers of infection**

To measure the number of infective centers, cells carrying PARIS or a control plasmid (pFR66) were grown in LB + Kan at 37°C to OD ~ 0.4. Cells were infected with T7 phage at a multiplicity of ~0.01 and incubated at 37°C for 5 min. Cells were then harvested (6,000 g – 3 min) and resuspended in LB + Kan to eliminate free phage, and serial dilutions in 100 μL of LB + Kan were prepared. To each dilution, we added 100 μL of phage-sensitive cells (MG1655 + pFR66), 5 mL of LB + 0.5% agar and CaCl₂ (5mM final concentration) and the mix was poured onto a LB + Kan plate. Care was taken to make sure the experiment was finished before lysis of infected cells (~20 min post-infection). The next day, infective centers were measured as the number of plaque-forming units on each plate.

**Isolation and sequencing of mutant phages**

We isolated T7 mutants that overcome PARIS by picking plaques in the lowest dilutions of a spot assay on a lawn of cells carrying PARIS from _E. coli_ B185. The resistance phenotype of each mutant was then verified by comparing the number of plaques in the presence or absence of PARIS. To extract phage DNA, 500 μL of high titer stocks (~10¹⁰ pfu/mL) of wild-type or mutant T7 were treated with TURBO DNase (Thermo Fisher Scientific) for 30 min at 37°C. DNase was inactivated with the addition of 5 μL of EDTA 0.5 mM and 25 μL of inactivation reagent at 65°C for 10 min. The supernatant was then treated with 0.5 mg/mL of proteinase K (Eurobio) and SDS 0.5% to release phage DNA from capsids. DNA was then purified as follows: 500 μL of a phenol-chloroform-isooamylalcohol (PCI, 25:24:1) solution (Sigma) were added to the sample which was then vortexed and centrifuged (6,000 g – 5 min). About 500 μL of the upper aqueous phase was transferred to a fresh tube and another 500 μL of PCI solution was added. After vortexing and centrifugation (6,000 g – 5 min), the upper aqueous phase was transferred to a tube containing 500 μL of chloroform. The sample was further vortexed and centrifuged (6,000 g – 5 min), and the upper aqueous phase was transferred to a tube containing 500 μL of cold isopropanol and incubated for 2h at -20°C to precipitate DNA. After centrifugation (16,000 g – 1 min), the DNA pellet was washed with 75% ethanol. Finally, the pellet was air-dried and resuspended in 50 μL of distilled water. Next-generation sequencing was performed using a Nextera XT DNA library preparation kit and the NextSeq 500 sequencing systems (Illumina) at the Mutualized Platform for Microbiology (P2M) at Institut Pasteur. Mutations were identified by mapping sequencing reads to the T7 reference genome using bzeroQ (v. 0.33.2) (Deatherage and Barrick, 2014). Additional T7 mutants...
were isolated from *E. coli* MG1655 + PARIS and *E. coli* DH10B + PARIS, the ocr gene was amplified using primers 5'-GTACGATG TACCACATGAAACG-3' and 5'-CACTCAAGCATATTCTAAAGCTTGTG-3' followed by Sanger sequencing (Table S4B).

**Transformation assays**

To verify the involvement of Ocr in PARIS, we cloned either a GFP or the T7 Ocr protein on the pBAD18 vector as follows: using Phusion PCR (ThermoFischer), a GFP fragment was amplified using primers 5'-ACCCTGTGTGTTTGGCTAGCGAATTGATATCCGGAGGCATATCAA-3' and 5'-GCCTTTCGTTTTATTTGATGCCTGGTTACTCTTAC-3' while ocr follows: using Phusion PCR (ThermoFischer), a GFP fragment was amplified using primers 5'-ACCCGTTTTTTTGGGCTAGCGAATTGCAAGGTGCCCTTTATGTATTA-3'. The pBAD18 backbone was amplified using primers 5'-ATTCCGCTAGCCCAAAAAAACGG-3' and 5'-CCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAATG-3'. Intact clones were inserted into the backbone using the Gibson method (Gibson et al., 2009). Constructions were electroporated into K-12 MG1655 cells and validated by Sanger sequencing, before plasmid extraction by miniprep (Macherey-Nagel). Electocompetent cells of MG1655 carrying PARIS from *E. coli* B185 or a control plasmid were prepared as described above. A hundred nanograms of plasmids were electroporated in both cell types and incubated for 1h at 37°C (induction with 0.3% arabinose) using a rubidium chloride chemical transformation protocol (Green and Rogers, 2013).

**CRISPRi-mediated EcoKI knockdown**

A sgRNA targeting hsdR (5'-ATGTTTTCCGGTGGGCCATT-3') was cloned onto plasmid psgRNAc (Addgene #114006) using the Golden Gate assembly method (Engler et al., 2008), giving pFD236, and introduced into the dCas9-expressing *E. coli* strain ACE1 (Calvo-Villamañán et al., 2020). The fitness of T7 or T7 mutants was then measured in the presence or absence of aTc (0.5 µg/ml), the inducer for dCas9 expression (Figure 2E).

**Effect of TIR-NLR on P2-P4 interaction**

*E. coli* C cells carrying either the control plasmid pFD232 or expressing the TIR-NLR system under the control of its native promoter on a low copy vector (pFD233), together with P420 were grown until OD~0.5 in LB + Cm + Kan and infected with P2 at an MOI of ~0.01. The culture was grown for 2 h at 37°C, centrifuged (5 min - 3,000 g) and the supernatant was recovered and filtered. To measure P420 titers, serial dilutions were prepared and 10 µl was mixed with 90 µl of *E. coli* C grown to OD~0.8 followed by incubation at room temperature for 30 min and plating on LB + Kan. Colony forming units were counted after overnight incubation at 37°C.

**P4-Kan construction**

P4-Kan is a variant of P4 sid1 (ATCC29746-B1) in which a Kan cassette was introduced in place of the *gpo* gene. It was constructed by 3-piece Gibson assembly. The first fragment containing the Kanamycin cassette was made using primers 5'-AATTTCGTGGCATGAGAGGATGAGAGATATCAG-3' and 5'-GGATAAGTATGTAATTCTTCAATAATACCTTCATGACGGCAACAAGATGAAAGTGCAGTTCACT-3' and 5'-GGATAAGTATGTAATTCTTCAATAATACCTTCATGACGGCAACAAGATGAAAGTGCAGTTCACT-3' from the same ATCC source. The second fragment containing the P4 satellite was amplified using 5'-GCTCATGTTGGTCAAGCTCGTGTTCAAC-3' and 5'-CCTTAACTCT-CTGACTGCACGGCAACAAGATGAAAGTGCAGTTCACT-3' from P4 sid1 (ATCC29746-B1) and the third fragment containing the other half of the P4 satellite was amplified using 5'-CTGAAACGGCTGCACAGAGAC-3' and 5'-ATGTTTTCCGGTGGGCCATT-3' from the same ATCC source. The assembly was transformed into DH10B cells and selected on LB + 50 µg/ml kanamycin plates.

**Competition between P2 and LF82_P8 in the presence of PARIS**

*E. coli* C cells carrying P4-Kan together with either the PARIS system under the control of its natural promoter on a low copy number vector (pFD235) or the control plasmid pFD232, were grown to OD~0.5 and infected with either P2, LF82_P8 or a 1:1 mixture of P2 + LF82_P8 at an MOI of ~0.01. The cultures were incubated for 2h30 at 37°C, centrifuged (5 min - 3,000 g) and the supernatant was recovered and filtered. Titers of P2 were determined by counting plaque forming units in a top-agar overlay of *E. coli* C carrying pFD235. The control experiment with LF82_P8 alone helped ensure that plaques measured in this manner were indeed P2 phage and not mutants of LF82_P8 that escaped defense by PARIS. Titers of LF82_P8 were determined by counting plaque forming units in a top-agar overlay of *E. coli* C lysogenized by P2. The control experiment with P2 alone helped ensure that plaques measured in this manner were indeed LF82_P8 phage and not mutants of P2 that escaped immunity provided by the P2 prophage. In parallel, serial dilutions of the infections were spotted on LB + Cm + Kan agar plates to estimate the number of live bacteria in the culture. The rate of lysisogeny during infection with P2 or P2+LF82_P8 in the presence of PARIS was determined by performing a P2-specific PCR on 16 isolated clones for each condition and each replicate of the experiment, using primers 5'-GTGTTTTAGTTACTAGATTGACG TACCTTATAG-3' and 5'-GGAAGATGAACCGTGAATGAGATATCCAG-3' (Figure S6).
**QUANTIFICATION AND STATISTICAL ANALYSIS**

**Identification of prophage-encoded systems**

We downloaded all 20,125 *E. coli* genomes and encoded protein sequences available on Genbank in August 2020 with any assembly status. We used the presence of the *psu* gene to pinpoint P4-like elements. To extract the systems located between *psu* and *int* genes in these elements, we performed a blastp search (version 2.9.0) against all *E. coli* proteins with an E-value threshold of $10^{-10}$ using the *Psu* protein sequence from *E. coli* E101 as a query (WP_000446153.1). For each hit, we searched for the presence of an integrase by keyword ("integrase") in the 10 downstream genes and retrieved the genes located in between, yielding a total of 5,251 loci after discarding loci that were across two contigs (Table S1 A). To reconstruct systems and assess their frequency, all 10,860 proteins from all loci were then clustered with MMseqs2 (version 8ebc9d16b2679eb485803259c8280127801e074b) (Steinegger and Söding, 2017) using a coverage threshold of 60% (-c 0.6 option), resulting in 541 clusters. For each locus, we then defined protein arrangements as a suite of clusters. We identified a total of 318 different protein arrangements. We manually curated the 121 protein arrangements that were present at least 5 times (together accounting for 94.4% of all loci) and grouped together highly similar systems sharing a core set of genes, resulting in 79 systems summarized in Table S1 B. To extract the systems located between *gpA* and *gpQ* genes in P2-like prophages, we performed a blastp search against all *E. coli* proteins with an E-value threshold of $10^{-10}$ using the *gpQ* protein from Enterophage P2 as a query (NP_046757.1). Similarly as described above, we searched for the presence of *gpA* by keyword ("replication endonuclease") in the 10 downstream genes and retrieved the genes located in between as candidate loci, yielding a total of 18,150 loci after discarding loci that were across two contigs. All 66,498 proteins from all loci were then clustered with MMseqs2 (-c 0.6 option), resulting in 846 protein clusters, and protein arrangements were reconstructed as described above, to the exception that we also discarded arrangements that only comprised proteins shorter than 120 amino acids. We identified a total of 1650 different arrangements, including 310 present in at least 10 occurrences (together accounting for 82% of all loci) that were manually curated as described above to yield 169 different systems listed in Table S5 B. Selected loci were visualized using clinker and clustermap.js (Gilchrist and Chooi, 2021).

Protein domains were identified using HHpred (Zimmermann et al., 2018) and HH-suite3 (Table S3; Steinegger et al., 2019). Transmembrane domains and signal peptides were identified using Phobius (Käll et al., 2007).

**Detection of PARIS**

In order to detect two-protein occurrences of the PARIS system (ATPase + DUF4435), we downloaded HMM profiles for the AAA_15 (PF13175), AAA_21 (PF13304) and DUF4435 (PF14491) protein families from the pfam database (El-Gebali et al., 2019). We then used these profiles to detect PARIS with MacSyFinder (v. 1.0.2) (Abby et al., 2014), requiring the two genes to be present (either AAA domain AND DUF4435), with the following score thresholds: 32 (AAA_15), 27 (AAA_21) and 27 (DUF4435) To detect single protein fusions, we browsed the “domain organisation” page of the DUF4435 pfam to look for proteins having both AAA_15/ AAA_21 and DUF4435 domains. We selected 10 protein sequences with AAA_15+DUF4435 domains and 10 protein sequences with AAA_21+DUF4435 domains and built HMM profiles from these. We then used this profile to detect PARIS with MacSyFinder (v. 1.0.2) (Abby et al., 2014) with the parameter “loner”, using the following score thresholds: 80 (AAA_15+DUF4435) and 20 (AAA_21+DUF4435). Using these detection rules, we analyzed 21,738 complete genomes retrieved from NCBI RefSeq in May 2021, representing 21 364 and 374 genomes of Bacteria and Archaea respectively.

**Identification of hotspots in other bacterial species**

From the IMG database (Chen et al., 2019) (accessed on December 1st 2020), we ran the built-in blastp function (e-value $10^{-10}$) using the DUF2290 and DUF4435 proteins from *K. pneumoniae* SB5961 and *E. coli* B185 respectively as queries. We then inspected the genomic neighborhood of distant hits using the online interface. We identified candidate hotspots when a neighboring gene frequently occurred next to the DUF2290 or DUF4435 proteins. In this way, we identified the *cl* gene from *Vibrionales* shown in Figure 7 A as well as the peptidase gene from *Baciliales* shown in Figure 7 B as potential locations for hotspots. To verify this, we then performed a new blastp search using these genes as queries against the IMG database and inspected the genomic neighborhood of their loci.
Supplemental information

Phages and their satellites encode hotspots of antiviral systems

François Rousset, Florence Depardieu, Solange Miele, Julien Dowding, Anne-Laure Laval, Erica Lieberman, Daniel Garry, Eduardo P.C. Rocha, Aude Bernheim, and David Bikard
Figure S1. Genomic characteristics of P4 and P2-encoded hotspots, Related to Figure 1 and Figure 4. Plots show the evolution of the frequency of each gene arrangement from P4-encoded (A) or P2-encoded (B) hotspots. Dashed lines highlight the analyzed threshold: (A) 121 arrangements occurring at least 5 times in the P4-encoded hotspot account for 94.4% of all P4-encoded loci, while (B) 316 arrangements occurring at least 10 times in the P2-encoded hotspot account for 82% of all P2-encoded loci. (C) We measured the size of the P4-encoded hotspot at each locus as the genomic distance between the end of psu and the end of int. For the P2-encoded hotspot, we used the distance between the end of A and the end of Q. P2-like phages encode larger systems than P4-like satellites on average (Mann-Whitney p-value < 10^{-100}).
Figure S2. Detailed fold-resistance of all verified defense systems against 8 phages, Related to Figure 2 and Figure 5. Plaque-forming units were measured for each of 8 phages on cells harboring either a control plasmid or a defense system from P4 (A) or P2 (B) hotspot. Fold-resistance was measured as the ratio between these two values. Bar plots show the mean of 3 to 4 independent measurements. All systems were expressed from their native promoter and measured at 37°C, except the RT-nitrilase + 1TM system that was measured at room temperature and PARIS-1 that was measured in the presence of aTc to induce expression. RT: reverse-transcriptase, TM; transmembrane helix, DUF: domain of unknown function, HP: hypothetical protein, TIR: Toll/Interleukin-1 Receptor, NLR: NOD-like Receptor, PRTase: phosphoribosyltransferase.
Figure S3. Genetic diversity encoded on P4-like satellites outside *E. coli*, Related to Figure 1. Genomic comparison of P4-like satellites in six different genomes of Enterobacteriaceae. P4 core genes are shown in blue shades. Grey shades show the percentage of identity between homologous proteins from different genomes. Genome accession numbers and positions are shown on the left.
Figure S4. Detection of PARIS in different prokaryotic phyla and defense phenotype of catalytic site mutants. Related to Figure 3. (A) PARIS was searched in 21,738 prokaryotic genomes using HMM searches and MacSyFinder (Tesson et al., 2021). For each clade, cells show the number and proportion of genomes where we detected PARIS. A color gradient is used to depict prevalence. (B) Genomic view of a few occurrences of PARIS in bacterial genomes. PARIS occurs either as a two-protein system or as a single protein fusion. PARIS genes are colored in red. Genome accessions are shown in parentheses and the bottom track shows genome coordinates. (C) Catalytic residues of PARIS in AriA and AriB were predicted using HHpred (Zimmermann et al., 2018). Resistance against phage T7 was measured for each gene deletion (D) and single or multiple amino acid mutants (E). Plaque assays are representative of three independent replicates. See also Table S7B.
Figure S5. Genomic comparison of canonical phage P2 with newly isolated P2-like phages, Related to Figure 5. Genomes were aligned and visualized using clinker and clustermap.js (Gilchrist and Chooi, 2021). Genes are colored by homologous groups. Grey shades illustrate protein identity.
Figure S6. Surviving colonies after co-infection of PARIS-expressing cells by P2 and LF82_P8 include P2 lysogens, Related to Figure 6. E. coli C carrying PARIS (pFD235) and P4-Kan was infected by P2 and LF82_P8 in 1:1 ratio (left) or by P2 alone as a control (right) (see methods). Clones obtained on LB + Cm + Kan plates after infection were screened for the presence of P2 lysogens. A P2-specific PCR was performed on 16 clones for each replicate. The number of positives was 2, 3 and 2 respectively after coinfection, and 9, 5 and 13 after infection by P2 alone. The estimated rate of lysogeny is thus ~15% after co-infection and 56% after infection by P2 alone. A P2 lysogen in E. coli C was used as a positive control for PCR (see methods). Representative agarose gels for each replicate are shown.
Figure S7. Genomic view of P4-like elements integrated near defense systems and integrative conjugative elements, Related to Figure 1. P4 genes are colored in blue while defense-associated genes are colored in orange or red. Genome accession numbers are shown in parentheses and the bottom tracks shows genomic coordinates.