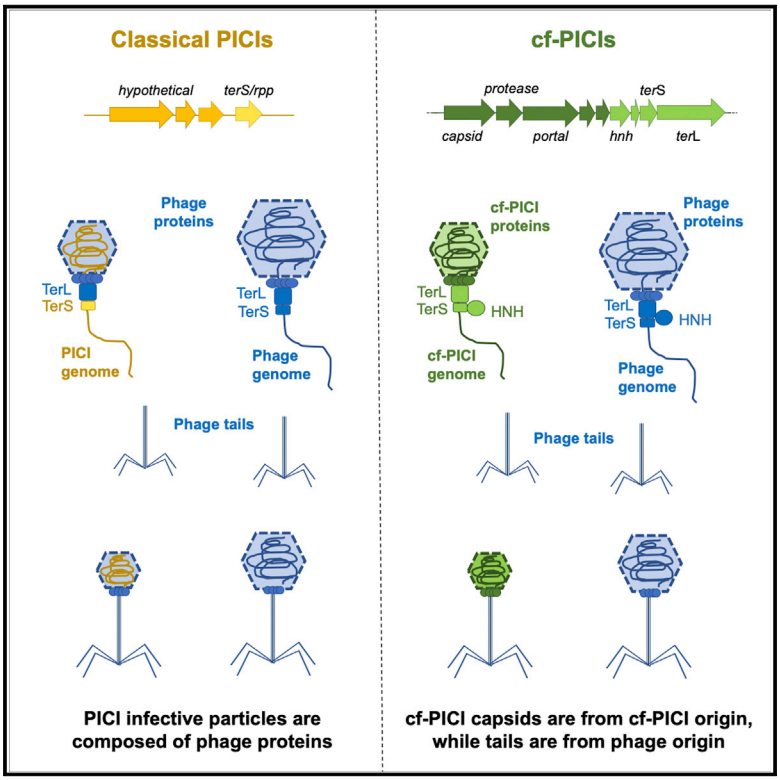
Article

Cell Host & Microbe

A widespread family of phage-inducible chromosomal islands only steals bacteriophage tails to spread in nature

Graphical abstract



Highlights

- Capsid-forming PICIs (cf-PICIs) represent a widespread family of phage satellites
- cf-PICIs produce small capsids to exclusively package their own dsDNA
- cf-PICIs only hijack phage tails to produce infective particles
- There are at least three groups of cf-PICIs that have emerged independently

Authors

Nasser Alqurainy, Laura Miguel-Romero, Jorge Moura de Sousa, John Chen, Eduardo P.C. Rocha, Alfred Fillol-Salom, José R. Penadés

Correspondence

a.fillol-salom@imperial.ac.uk (A.F.-S.), j.penades@imperial.ac.uk (J.R.P.)

In brief

Alqurainy et al. discovered a widespread family of phage satellites encoding all the proteins required to produce small capsids, where satellite genomes are exclusively packaged. Therefore, to produce infective particles, these satellites only parasitize phage tails. These findings represent a new type of interaction between satellites and helper phages.

Alqurainy et al., 2023, Cell Host & Microbe *31*, 69–82
January 11, 2023
https://doi.org/10.1016/j.chom.2022.12.001

CellPress

Cell Host & Microbe

CellPress
OPEN ACCESS

Article

A widespread family of phage-inducible chromosomal islands only steals bacteriophage tails to spread in nature

Nasser Alqurainy,[1,2,7] Laura Miguel-Romero,[1,3,7] Jorge Moura de Sousa,[4] John Chen,[5] Eduardo P.C. Rocha,[4] Alfred Fillol-Salom,[1,3,*] and José R. Penadés[3,6,8,*]

[1]Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8TA, UK

[2]Department of Basic Science, College of Science and Health Professions, King Saud bin Abdulaziz University for Health Sciences, Riyadh 11426, Saudi Arabia

[3]Centre for Bacterial Resistance Biology, Imperial College London, London SW7 2AZ, UK

[4]Institut Pasteur, Université de Paris Cité, CNRS, UMR3525, Microbial Evolutionary Genomics, 75015 Paris, France

[5]Department of Microbiology and Immunology, Infectious Diseases Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore

[6]Universidad CEU Cardenal Herrera, CEU Universities, Valencia 46115, Spain

[7]These authors contributed equally

[8]Lead contact

*Correspondence: a.fillol-salom@imperial.ac.uk (A.F.-S.), j.penades@imperial.ac.uk (J.R.P.)

https://doi.org/10.1016/j.chom.2022.12.001

SUMMARY

Phage satellites are genetic elements that couple their life cycle to that of helper phages they parasitize, interfering with phage packaging through the production of small capsids, where only satellites are packaged. So far, in all analyzed systems, the satellite-sized capsids are composed of phage proteins. Here, we report that a family of phage-inducible chromosomal islands (PICIs), a type of satellites, encodes all the proteins required for both the production of small-sized capsids and the exclusive packaging of the PICIs into these capsids. Therefore, this new family, named capsid-forming PICIs (cf-PICIs), only requires phage tails to generate PICI particles. Remarkably, the representative cf-PICIs are produced with no cost from their helper phages, suggesting that the relationship between these elements is not parasitic. Finally, our phylogenomic studies indicate that cf-PICIs are present both in gram-positive and gram-negative bacteria and have evolved at least three times independently to spread in nature.

INTRODUCTION

Mobile genetic elements (MGEs) are key players driving bacterial evolution and ecology.[1] Phage satellites are an important type of MGEs that parasite phages to ensure their promiscuous dissemination in nature, at the expense of their inducing helper phages. To date, the most studied phage satellites are P4 in *Enterobacterales*,[2] phage-inducible chromosomal islands (PICIs) in *Bacillales* and *Gammaproteobacteria*,[3,4] and PICI-like elements (PLEs) in *Vibrio* spp.,[5,6] with thousands of these elements widely distributed in hundreds of different bacterial species.[7] Although the genetic organization of these families of satellites is different, all these elements encode genes required for their regulation, replication, and preferential packaging at the expense of their helper phages.[2,5,8,9] Importantly, these elements do not encode the proteins required for the formation of their infective particles, and therefore, for their transmission, they must hijack the structural proteins and assembly processes of their helper phages.[8,10]

Satellites have evolved different strategies to promote their preferential packaging into capsids composed of phage-encoded proteins. A strategy commonly used by satellite phages to interfere with helper phage packaging is the production of small capsids, which are commensurate to the size of the satellite's genomes but too small to accommodate the complete helper phage genome. Remarkably, although this strategy is conserved, satellites use different mechanisms to redirect phage capsid assembly, which represents a nice example of convergent evolution. While the satellite phage P4 expresses Sid (an external scaffolding protein) to redirect P2 capsid assembly,[11] PICIs have evolved multiple mechanisms, depending on the phage they parasite (recently reviewed by Fillol-Salom et al.[12]). Thus, some staphylococcal pathogenicity islands (SaPIs), which are the prototypical members of the PICI family, express CpmB, which operates as a scaffolding protein, binding to the phage capsid protein, altering the curvature of the shells and redirecting its assembly and capsid size. This system also requires CpmA, whose predicted function is to remove the phage scaffolding

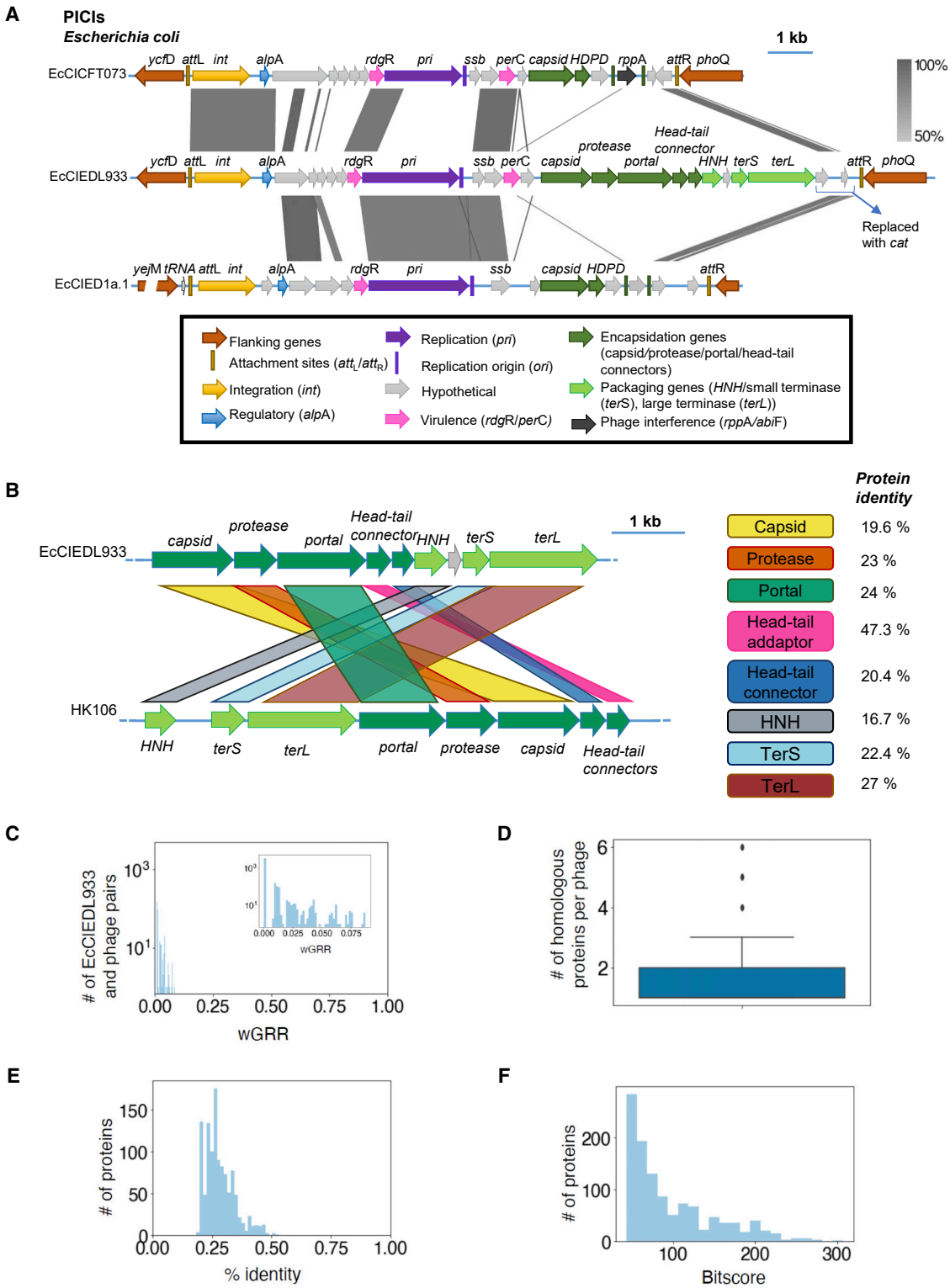


Figure 1. EcCIEDL933 represents the prototypical member of a new family of PICI elements

(A) Comparative maps of the classical *E. coli* PICIs EcCICFT073 and EcCIED1a.1 with EcCIEDL933. Genomes are aligned according to the prophage convention, with the integrase genes (*int*) present at the left end of the islands. Genes are colored according to their sequence and function (see box for more details). Using Easyfig alignment, vertical blocks between PICI sequences indicate regions that share similarity based on BLASTn (gray scale).

(B) Comparison of the packaging modules present in either EcCIEDL933 or in helper phage HK106, at protein level, using BLASTp. Protein identities obtained with Clustal Omega are shown for each protein.

(legend continued on next page)

protein.[9,13,14] Other SaPIs that use the *cos* system for packaging express Ccm, which has homology with the helper phage major capsid protein and redirects the assembly of the phage-encoded capsid protein into the production of SaPI-sized small capsids.[15–17] PICIs from *Enterococcus faecalis* (EfCIV583)[18] and *Pasteurella multocida* (PmCI172)[3] and PLEs from *Vibrio cholerae*[19] have also the ability to produce satellite-sized capsids; although in these cases, the proteins involved in this process have not been identified yet.

While capsid redirection severely reduces helper phage reproduction, it does not increase *per se* the packaging of the satellite phages into the small capsids. To solve this problem, satellites have evolved an arsenal of complementary and sophisticated strategies that ensure their preferential packaging into the virions, facilitating their promiscuous transfer in nature. Phage packaging starts by the recognition of the *cos* or *pac* sites present in the phage genomes by the phage-encoded small terminase (TerS). To promote their packaging, some satellites—such as P4 and some SaPIs—carry the cognate helper phage *cos* site in their genomes.[15,20] Other SaPIs encode a homolog of the phage TerS called $TerS_S$,[14] which recognizes the SaPI *pac* sequence.[21] To help $TerS_S$ function, these SaPIs also encode phage packaging interference (Ppi), which binds to the phage TerS, blocking its function.[22,23] Finally, some PICIs present in *Proteobacteria* encode redirecting phage packaging (Rpp), which binds to the phage TerS. The Rpp-TerS complex recognizes the PICI *cos* site (promoting PICI packaging) but not the phage *cos* site, thus blocking phage reproduction.

While diverse, all the mechanisms of piracy described so far require that helper phages express all the proteins necessary to generate the PICI infective particles. We report here the discovery of a widespread family of PICIs that we have named as capsid-forming PICIs (cf-PICIs) that encode all the proteins required not only for the production of the PICI-sized capsids but also for the exclusive packaging of the cf-PICIs into the small-sized capsids. Therefore, contrary to what is seen with other satellites, cf-PICIs only require phage tails to spread in nature. They represent a new strategy of molecular piracy.

RESULTS

EcCIEDL933 is the prototypical member of the cf-PICI family

During the description of PICI elements in *Proteobacteria*,[3] one intriguing element with an unusual genetic organization, present in the *Escherichia coli* strain EDL933, raised our curiosity. This element, with a size of 15.4 kb, carried the *int* and *alp*A genes identical to those present in the prototypical *E. coli* EcCICFT073 PICI, while the *pri* gene was identical to that present in EcCIED1a.1 (Figure 1A). However, in addition to these conserved PICI genes, this element seemed to encode all the proteins required to produce functional capsids (major capsid, head maturation protease [HMP], portal, and head-tail connectors) and for the packaging of the PICI DNA into these capsids (HNH endonuclease and small (TerS) and large (TerL) terminase proteins) (Figure 1A). Interestingly, this element did not encode genes involved in the production of phage tails.

Based on this peculiar organization, our initial thoughts were that this element could represent a hybrid between a PICI and a phage. Note that the aforementioned DNA packaging genes present in this putative hybrid element showed homology with phage genes (Figure 1B). Yet, sequence similarities between the genes present in this new element and those present in phages were generally low and restricted to a few genes. To assess the similarity in gene repertoires between the element present in strain EDL933, hereinafter named EcCIEDL933, and the genomes of typical phages, we computed their weighted gene repertoire relatedness (wGRR), which we have used previously to class phages in families.[25] wGRR quantifies the similarity between two genomes based on the frequency of bi-directional best protein matches weighted by the protein sequence identity. It varies between 0 and 1. It has a value of 0 when the elements lack homologs. It has a value of 1 when all the genes in one element have a bi-directional best hit in the other, and these homologs are 100% identical in protein sequences (see STAR Methods). Our analysis of 3,725 complete non-redundant phage genomes retrieved from RefSeq shows that the genome of EcCIEDL933 is very different from them. Indeed, 86% of the phage genomes lack homologs in EcCIEDL933, with the remaining 14% (519 phage genomes) having, at most, a wGRR of 0.09 (Figure 1C). These latter cases represent phages with a few homologs with low sequence identity to the proteins encoded by EcCIEDL933. They include, besides the aforementioned capsid and terminase genes, homologs to the other capsid formation (portal, head-tail connectors) or packaging (HNH) genes. Importantly, no single phage genome has more than six homologous proteins with EcCIEDL933, and the average number of homologs, among phages with at least one homolog, is only 1.8 (Figure 1D). Most of these rare homologs have significant but low sequence similarity (<40%, e value < 1e−4; Figures 1E and 1F). On the other hand, our Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology analyses showed that genes homologous to the ones present in EcCIEDL933 are frequently found in genomic regions that could correspond to similar elements in *E. coli* and other related species, including *Salmonella enterica*, *Shigella flexneri*, *Klebsiella pneumoniae*, *Pluralibacter gergoviae*, or *Yersinia aleksiciae* (Table S1). Therefore, these results suggested that EcCIEDL933 could represent a new family of elements, distinct from both phages and previously described PICIs. In support of this idea, members of this new family integrate into the *E. coli* bacterial chromosome using six different *att*B sites. Of these, two were exclusively used by the members of this new lineage (*gcd-hpt*A and *yhd*J-*fis*), while the other four were also

(C) Distribution of wGRR values between EcCIEDL933 and phage genomes. Inset shows a zoomed version of wGRR between 0 and 0.1.
(D) Boxplot with the number of homolog genes per phage for all the phages in the database.
(E) Distribution of the protein identity percentages for all the potential homologs between EcCIEDL933 and phage genomes.
(F) Distribution of the bitscores for all the potential homologs between EcCIEDL933 and phage genomes. See also Figure S1 and Tables S1 and S2.

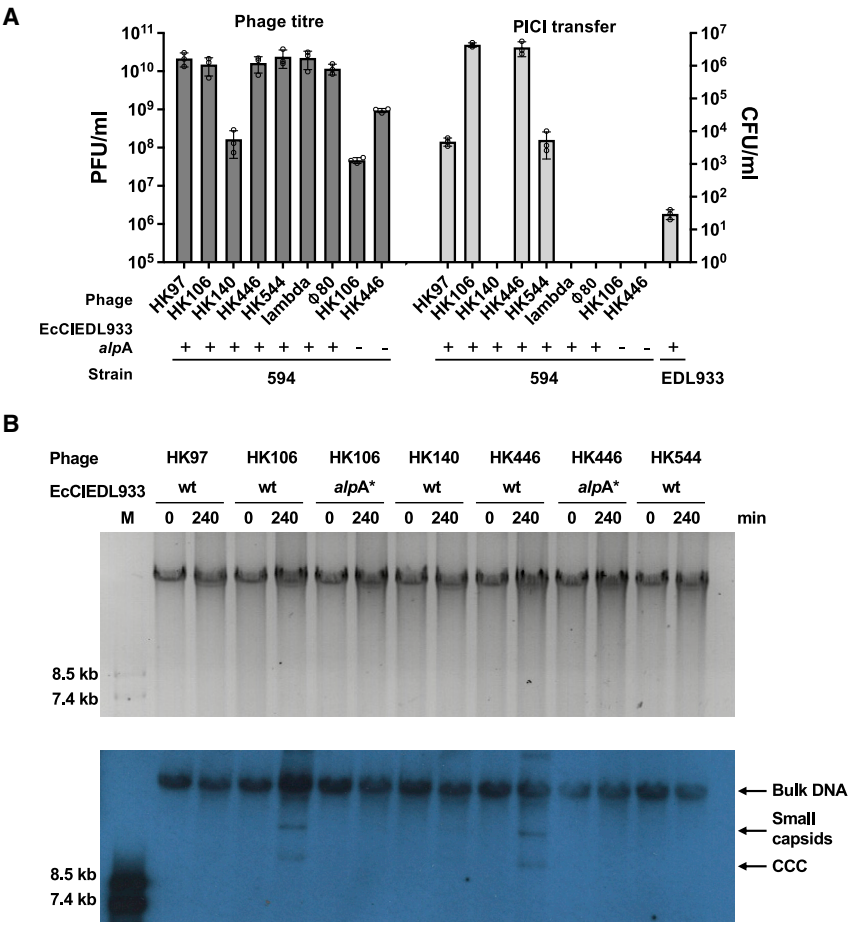


Figure 2. Phage induction of EcCIEDL933

(A) Different lysogenic strains carrying EcCIEDL933 (WT or mutant in *alp*A) were MC (2 μg/mL) induced, and the phage and PICI titers were determined using *E. coli* 594 as the recipient strain. The means of the phage-forming units (PFUs) or the colony-forming units (CFUs) and SD of three independent experiments are represented (n = 3). The limits of detection in these experiments are 10 PFUs or CFUs, respectively.

(B) Different lysogenic strains carrying EcCIEDL933 (WT or mutant in *alp*A [*alp*A*]) were MC (2 μg/mL) induced; samples were taken at the indicated time points (min) and processed to prepare minilysates, which were separated on a 0.7% agarose gel (upper panel, GelRed stain), and Southern blotted (lower panel) with an EcCIEDL933 probe. M stands for Southern blot molecular marker (DNA molecular weight marker VII; Roche). Bulk DNA is the concatemeric form. Small capsids correspond to the DNA packaged into the PICI-sized capsids. CCC is covalently closed circular molecules. See also Figure S2 and Table S3.

used by members of the previously identified *E. coli* PICIs (Table S2; Figure S1).

The fact that EcCIEDL933 encodes *int*, *alp*A, and *rep* genes almost identical (98%, 94%, and 94% identity, respectively) to those present in prototypical PICIs suggested that the mechanism of induction of this new lineage could be identical to that previously described for the PICIs present in *E. coli*. Once induced, the genes involved in DNA packaging could either promote the packaging of the cognate element and/or could be used by the island to interfere with helper phage reproduction. To clearly demonstrate that EcCIEDL933 is induced as a classical PICI element, we searched for helper phages that are able to induce and mobilize EcCIEDL933. To facilitate these studies, we inserted a chloramphenicol resistance marker (*cat*) into the EcCIEDL933 element present in the clinical EDL933 strain (Figure 1A), which would allow us to detect the island during its mobilization. This island, carrying the *cat* marker, was the only EcCIEDL933 derivative used through all the experiments; so, to facilitate the reading, we have just called it EcCIEDL933. Since the EDL933 strain also carries 17 prophages,[26] we induced the resident prophages with mitomycin C (MC), expecting that one of the resident prophages would mobilize EcCIEDL933 into the non-lysogen 594 *E. coli* strain, used in this experiment as a recipient. Only a few transductants were obtained (Figure 2A), suggesting that none of the resident prophages present in the EDL933 strain is a helper phage for this island. Note that in presence of a helper phage, PICI transfer is normally extremely high.[27] Indeed, Southern blot analyses confirmed that none of the 17 prophages induced EcCIEDL933 since PICI replication was not detected after MC induction of the EDL933 strain (Figure S2A). We ignore the reasons for this result, but we suspect that either the helper phage did not integrate into the EDL933 genome or it was inactivated or lost since then. This is not unusual. Previous work has shown that *E. coli* strains may carry many partially defective prophages.[28]

To search for a phage helper, we lysogenized the *E. coli* 594 strain carrying EcCIEDL933 with our collection of *E. coli* phages and tested the transfer of the PICI element after MC induction of different lysogenic strains. Two phages, HK106 and HK446 (see characteristics in Table S3), mobilized the island at high frequencies (Figure 2A), suggesting they were able to induce the EcCIEDL933 cycle. Note that as occurred with EcCICFT073, the prototypical member of the *E. coli* PICIs,[3] the ratio of PICI versus helper phage particles obtained after MC induction of these strains was 1:1,000, suggesting that these elements establish a different interaction with their helper phages in this species. To demonstrate that HK106 and HK446 are the helper phages for the EcCIEDL933 element, a screening lysate analysis of the DNA samples obtained after induction of the different lysogenic strains carrying EcCIEDL933 was performed. As shown in Figure 2B, only the strains lysogenic for either HK106 or HK446 showed the classical pattern of PICI induction: a large-sized DNA band corresponding either to the PICI or phage replicating concatemeric forms or to the PICI or phage DNA packaged into the phage-sized capsids ("bulk DNA" band) and a small band corresponding to the PICI DNA packaged into small PICI-sized capsids ("small capsids" band).[4,14]

Finally, we tested the transfer of the EcCIEDL933 *alp*A mutant. Note that in gram-negative bacteria, PICI induction and transfer requires the expression of the PICI-encoded *alp*A gene, a process that is activated by the helper phage.[3] As expected for a PICI, mutation of *alp*A eliminated both induction by helper phages (Figure 2B) and transfer of this element (Figure 2A).

EcCIEDL933 produces PICI-sized capsids

Having established that the prototypical member of this new family of satellites is induced like the classical PICIs, we wanted to know whether the EcCIEDL933 genes with homology to phage genes were required for capsid formation and EcCIEDL933 packaging and transfer. The presence of the PICI-sized small band in the EcCIEDL933 screening lysate suggested that this island may have the ability to produce PICI-sized capsids after induction. To test this, we analyzed by electron microscopy (EM) the infective particles present in the lysate obtained after induction of the lysogenic strain for helper HK106, in the presence or absence of EcCIEDL933. In the absence of EcCIEDL933, only phage particles were observed, which had the characteristic size and shape of a *Siphoviridae* phage with a hexagonal capsid of approximately 58–61-nm width, with non-contractile tails, which are approximately 162–168 nm long (Figure 3A; Table S4). By contrast, in the presence of EcCIEDL933, two different types of infective particles were observed: in addition to the phage-sized ones, we were able to detect PICI-sized capsids, which had a hexagonal capsid of 45–47-nm length, connected to the phage non-contractile tail (Figure 3A; Table S4).

Since the PICI genome is one-third of the size of the phage genome, it is obvious that the complete phage genome cannot be packaged into the PICI-sized particles. However, some PICIs can package three copies of their genomes into phage-sized capsids.[14,16] To test whether this was also the case for EcCIEDL933, the packaged DNA present in the lysate containing both the phage and the PICI infective particles was purified, separated on agarose gels, and analyzed by Southern blot analyses, using specific phage or PICI probes. In support of the EM images, two types of bands were observed, corresponding to the size of the phage and EcCIEDL933 genomes, respectively (Figure 3B). Importantly, the Southern blot analyses revealed that the packaging of the different elements is size specific and mutually exclusive, with phages and PICIs being packaged uniquely in large- and small-sized capsids, respectively (Figure 3C). These results suggest that EcCIEDL933 can produce small capsids that are composed exclusively of PICI proteins. It is tempting to speculate that the PICI-sized capsids have a $T = 4$ icosahedral structure.

Characterization of the EcCIEDL933 packaging module

The previous results suggested that the different sized capsids had different origins: the phage would be responsible for the production of large capsids and phage tails, while the small capsids would be produced by the PICI. In support of this idea, a comparison of the phage and PICI packaging modules revealed that they have different genetic organizations and low protein identities (Figure 1B). To test this idea, we individually introduced stop codons to mutate each of the genes belonging to the putative packaging module of EcCIEDL933 (Figure 4A) and tested the transfer of the different EcCIEDL933 mutants after induction of the helper prophage HK106. Remarkably, while the replication of the different mutants was unaffected (Figure S3), their transfer was completely abolished in all cases except in the mutant for the hypothetical gene 1792, whose transfer was severely affected (Figure 4B). To confirm that the observed phenotypes were consequence of the mutations, complementation of the mutants partially restored PICI titers (Figure S4C). In all these cases, the titer of the helper phage was unaffected, confirming that the PICI-encoded proteins have no role in the packaging of the helper phage (Figure 4C). Importantly, the Southern blot analyses of the purified infective particles obtained after induction of the different EcCIEDL933 mutant strains revealed the presence of only a phage-sized band but not PICI-sized ones, confirming that the small capsids had a PICI origin (Figure S5C).

Next, we performed the same experiment but now mutating the phage genes encoding the homolog proteins to those encoded by the PICI. In this case, the mutations eliminated phage but not PICI transfer (Figures 4E and 4F). Phage mutants were complemented in *trans*, confirming that the phenotypes were due to the mutations (Figure S4D). Moreover, in agreement with the idea that the large and small capsids had different origins, only PICI-sized bands were observed in the Southern blot analyses of the purified infective particles obtained after induction of the different phage HK106 mutant strains (Figure S5D). This was corroborated with the analysis of the lysate obtained after induction of the strain carrying the HK106 *ter*L (*gp01*) or *capsid (gp05)* mutation and EcCIEDL933, which only contained PICI-sized particles (Figure 3A). Indeed, the packaged DNA from HK106 *ter*L (*gp01*) mutation and EcCIEDL933 reveals only a small band on the agarose gel, which was analyzed by Southern blot analysis and corresponds uniquely to EcCIEDL933 (Figures 3B and 3C). Furthermore, to clearly confirm the origin of the PICI-sized capsids, the lysate obtained after induction of the strain HK106 *gp05** EcCIEDL933 was purified, the proteins were run on an SDS-PAGE gel, and the bands corresponding to the most abundant proteins were analyzed. In support of our hypothesis, the capsid and portal proteins obtained from the PICI infective particles were PICI encoded, while the major tail protein was of phage origin (Figure 5A; Table S5).

Finally, to clearly demonstrate that the phage HK106 is exclusively packaged in the large capsids (phage origin), while EcCIEDL933 is exclusively packaged in the small ones (PICI origin), we induced the strain carrying the wild type (WT) phage HK106 and the mutant EcCIEDL933 *1786** island (capsid mutant), which only produces large capsids (Figure 3A), as well as the strain carrying the mutant HK106 *gp05** (capsid) phage and the WT EcCIEDL93, which only produces PICI-sized capsids (Figure 3A). After induction, the infective particles were purified, and the packaged DNA was sequenced. In support of our hypothesis, the HK106 DNA was in majority in the large capsids obtained after induction of the strain carrying HK106 and EcCIEDL933 *1786**, while the EcCIEDL933 DNA was in majority in the small capsids obtained after induction of the strain carrying HK106 *gp05** EcCIEDL933 (Figure 5B). Note that both phage and PICI DNAs were detected when we repeated this experiment with strains carrying WT versions of the HK106 and EcCIEDL933 elements (Figure 5B), which produced both phage and EcCIEDL933 particles (Figure 3A). Although in these

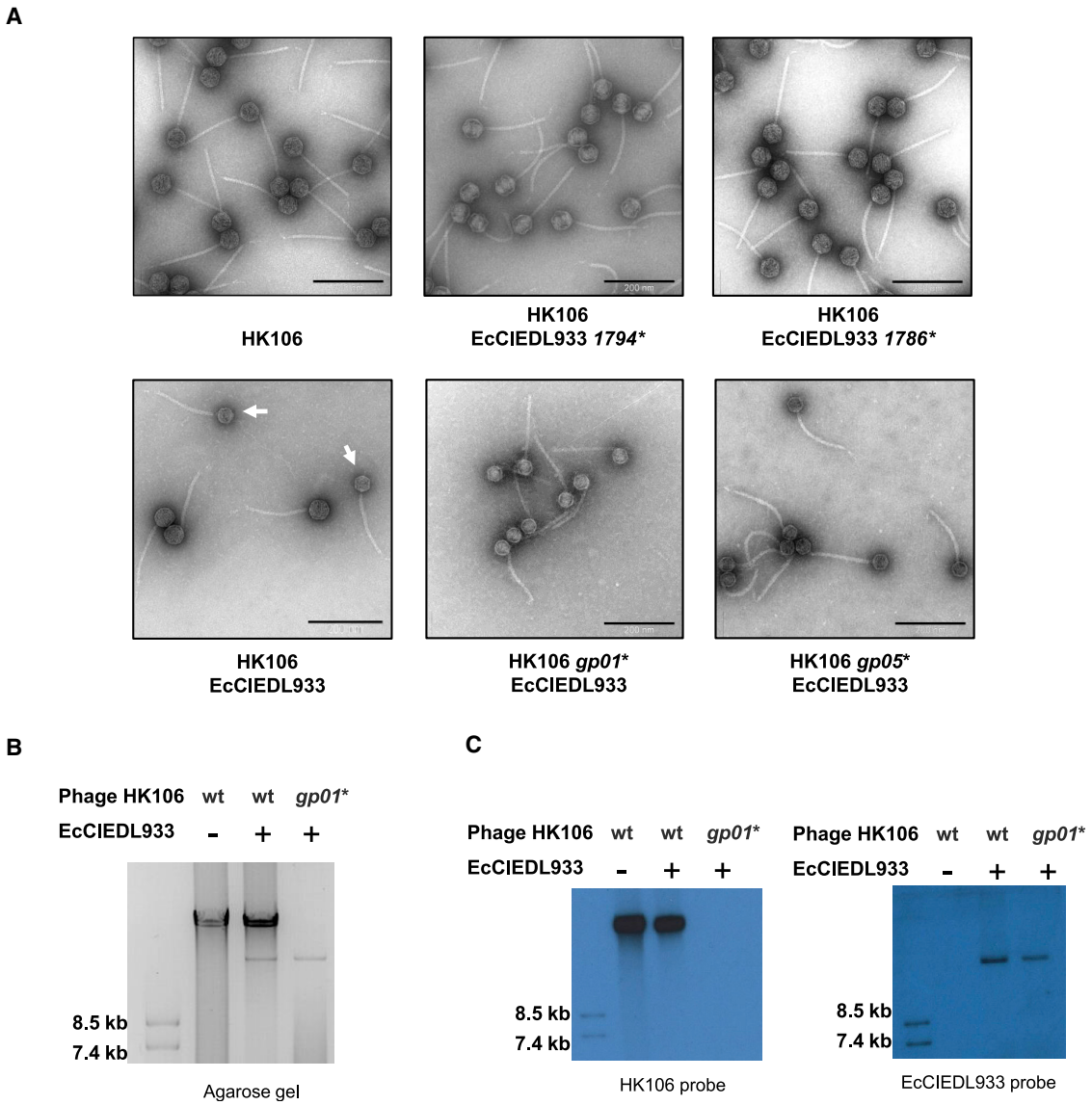


Figure 3. EcCIEDL933 produces small capsids

(A) Electron microscopy analyses of different MC-induced HK106 lysogenic strains (WT, *ter*L mutant [*gp01**] or capsid mutant [*gp05**]) in presence or absence of EcCIEDL933 (WT, *ter*L mutant [*1794**] or capsid mutant [*1786**]). Different fields are shown, containing HK106 phage particles (top images), EcCIEDL933 particles (bottom images), or both (bottom left; PICI particles highlighted with white arrows). EcCIEDL933 particles have smaller heads. Images were collected using a JEOL 1200 TEM microscope with 12K magnification. Scale bars represent 200 nm.

(B) Lysogenic strains carrying the WT or the *ter*L mutant (*gp01**) HK106 prophage, in presence or absence of EcCIEDL933, were MC induced; the DNA was extracted from the purified infective particles and resolved on a 0.7% agarose gel, GelRed stained. The size of phage HK106 is 41,468 bp, while EcCIEDL933 has a size of 15,471 bp.

(C) Southern blot of the purified DNA shown in (B) using either phage HK106- or EcCIEDL933-specific probes. First line contains the Southern blot molecular marker (DNA molecular weight marker VII; Roche). See also Table S4.

experiments the different samples were treated with DNase before and after CsCl density gradient purification, some chromosomal, phage, and PICI contaminations were observed, probably because the binding of the contaminant DNA to the infectives particles protects it against DNase degradation. In any case, it is clear from these studies that the preferential package of the phage and PICIs is present in large- and small-sized capsids, respectively.

Our previous results demonstrated that EcCIEDL933 encodes a functional packaging module responsible for both the production of the PICI-sized capsids and for the specific and exclusive packaging of the island into these particles. Therefore, we propose here that EcCIEDL933 is the prototypical member of a new family of PICI satellites, which we have named as capsid-forming PICIs (cf-PICIs), with the ability to produce all the proteins required for the specific packaging of these elements into

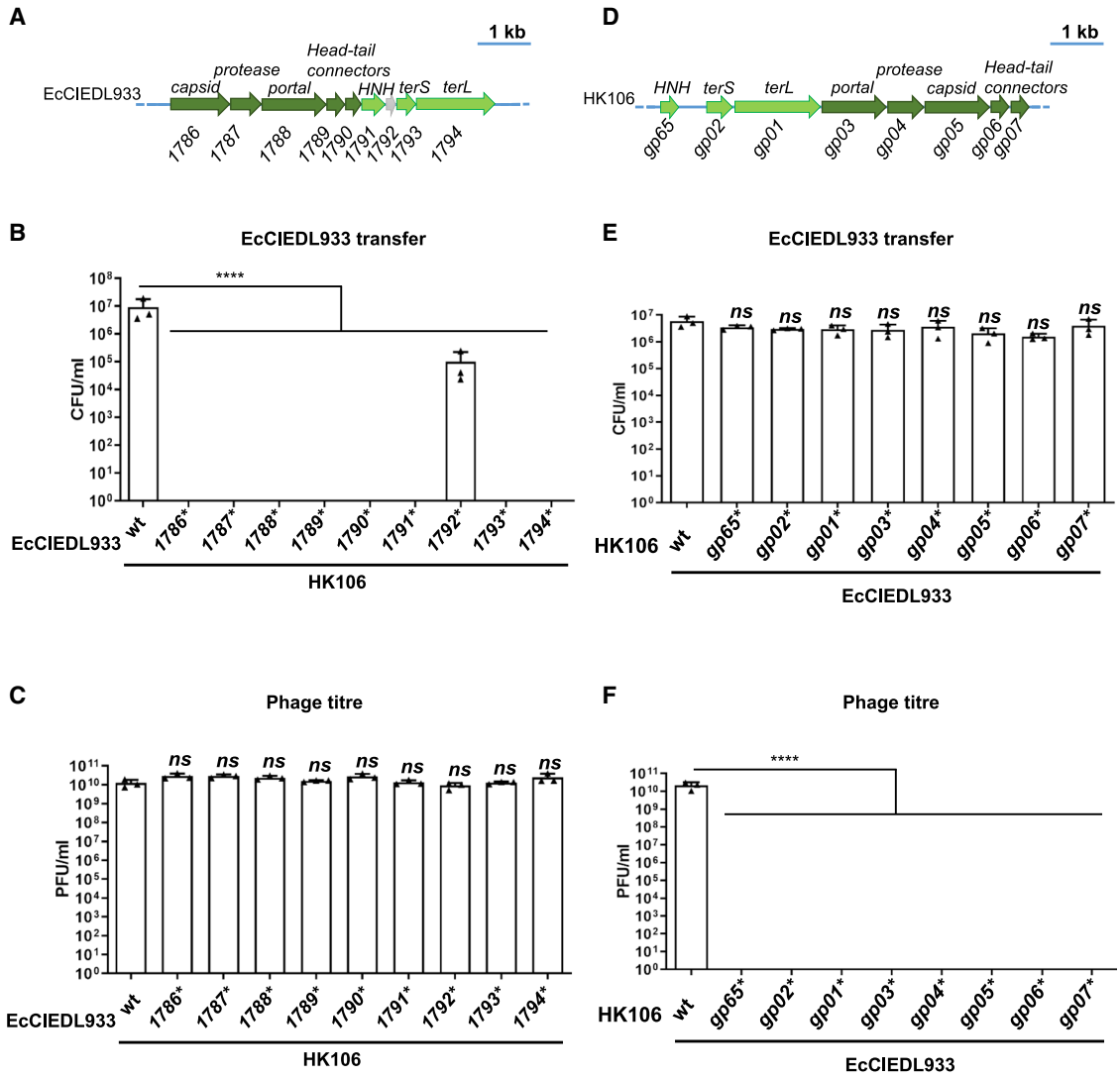


Figure 4. Effect of different PICI and phage gene mutations on EcCIEDL933 and HK106 transfers

Genetic maps of the EcCIEDL933 (A) or the phage HK106 (D) packaging modules. Lysogenic strains for phage HK106, carrying different EcCIEDL933 versions (WT or carrying mutations in the indicated genes), were MC (2 μg/mL) induced, and the PICI (B) or phage (C) transfers were determined using *E. coli* 594 as the recipient strain. Lysogenic strains for HK106, carrying the WT or the different mutant prophages, were MC (2 μg/mL) induced in the presence of EcCIEDL933, and the transfer of the island (E) or the phage titers (F) were determined using *E. coli* 594 as the recipient strain. The means of the colony-forming units (CFUs) or phage-forming units (PFUs) and SD of three independent experiments are presented (n = 3). The transduction and titration limit of detection is 10 CFUs or PFUs, respectively. A one-way ANOVA with Dunnett's multiple comparisons test was performed to compare mean differences between WT and individual mutants. Adjusted *p* values were as follows: ns > 0.05; *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001; ****p ≤ 0.0001. See also Figures S3–S5.

cf-PICI-encoded capsids. These results also suggested that in terms of producing infective particles, EcCIEDL933 only requires the phage tails to complete the formation of the PICI infective particles. To confirm this hypothesis, we analyzed the HK106 *gp05** EcCIEDL933 whole infective particle by mass spectrometry. Importantly, this approach confirmed that the only protein from phage origin was the phage-encoded major tail protein (Table S5). Additionally, we mutated two of the phage genes encoding important tail components: the tape measure (*gp14*) and major tail (*gp10*) proteins. The strains carrying the mutant phages and EcCIEDL933 were MC induced, and the phage and PICI titers were quantified. As expected, these mutants abolished both phage and PICI transfer (Figure 6A). Next, to confirm that EcCIEDL933 only relies on the phage tail proteins to produce PICI infectious particles, we analyzed by EM the lysates obtained after the induction of the strain containing only the HK106 prophage (WT or mutant in the major tail gene [*gp10*]) and those that contained the PICI and a helper prophage carrying mutations in the major tail (*gp10*) and the capsid (*gp05*) genes. Note that in the latter strain, the phage does not produce large capsids nor tails. In support of our hypothesis, only phage-sized (HK106 *gp10** strain) or PICI-sized (HK106 *gp05** *gp10** EcCIEDL933 strain) capsids were observed, respectively, in these lysates (Figure 6B).

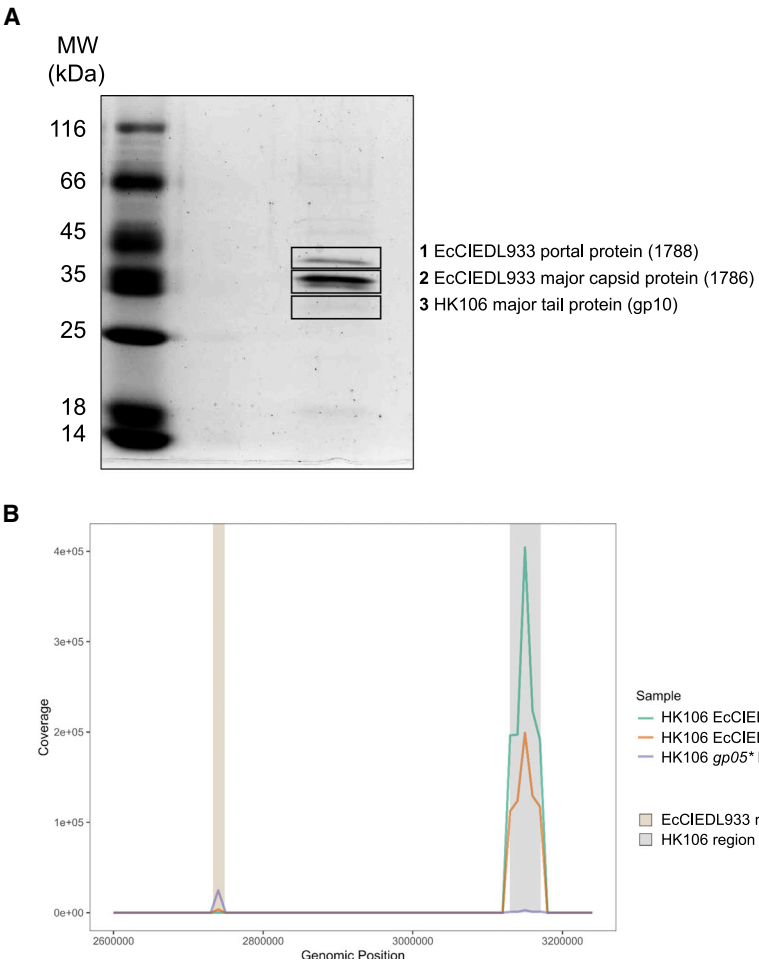


Figure 5. DNA sequencing and mass spectrometry of viral particles

(A) Purified HK106 *gp05** EcCIEDL933 particles were run in a 12% acrylamide/bis SDS-PAGE gel. The indicated bands 1, 2, and 3, highlighted with squares in the figure, were extracted and analyzed by mass spectrometry. The mass spectrometry results are collected in Table S5.

(B) Lysogenic strains for phage HK106 carrying either EcCIEDL933 (produces phage and PICI particles) or EcCIEDL933 *1786** (only produces phage particles), or a lysogenic strain for phage HK106 *gp05** carrying EcCIEDL933 (only produces PICI particles) were mitomycin C induced, and the infective particles were purified. The packaged DNA was then extracted, purified, and sequenced using Illumina NextSeq 2000 platform. The coverage of chromosomal DNA is represented using *E. coli* 594 strain as a template containing the integrated phage and PICI genomes. The genomic regions containing either EcCIEDL933 PICI or prophage HK106 are highlighted in brown and gray, respectively. See also Table S5.

EcCIEDL933 does not block helper phage reproduction

One conserved characteristic of the satellites is their ability to severely interfere with their helper phage reproduction because they hijack their components.[5,12] To test whether EcCIEDL933 was also able to block the reproduction of its helper phages, we quantified the phages obtained after induction of the lysogenic strains for HK106 or HK446, the EcCIEDL933 helper phages, in presence or absence of the island. Importantly, reproduction of phages HK106 or HK446 was unaffected by the presence of EcCIEDL933 (Table S6). To confirm this, we infected strains 594 or JP22295 (a 594 derivative carrying EcCIEDL933) with phages HK106 or HK446 and compared their efficiency of plating. With the classical PICIs, in presence of the islands, both the number and the size of the helper phage plaques are severely reduced since only evolved phages, which are unable to induce the island, are able to form plaques.[29] In this case, however, the scenario was completely different, and after phage infection, the presence of the island did not affect the capacity of the helper phages to reproduce normally in the recipient strains carrying the island, compared with the EcCIEDL933-negative strain (Figure S2B). These results suggest that EcCIEDL933 is not a phage parasite but rather a commensal under the conditions tested.

cf-PICIs are widespread in nature

We next analyzed whether this new family of satellite phages is widespread in nature. We curated a search across the GenBank database for similar cf-PICI elements among 13 species of *Proteobacteria*, especially members of *Enterobacteriaeae*, *Pasteurellaceae*, and *Orbaceae*, where PICIs are frequent.[3] We searched for elements carrying *int* or *alp*A genes identical to those present in the canonical PICIs but encoding a packaging module similar to that present in the cf-PICI EcCIEDL933. Our preliminary analysis revealed that other cf-PICIs are frequent in *Proteobacteria* genomes (Figure S1; Table S2). Importantly, all the elements identified encode all, or most, of the components present in EcCIEDL933, with a similar conserved genetic organization (Figures 1 and S1). Their genes encoding the packaging module present sequence similarity with those from EcCIEDL933 with a range of 50%–90%, suggesting that they are indeed part of a single family of satellites.

Using the same strategy that allowed us the identification of the cf-PICIs in *Proteobacteria*, we next interrogated the genomes of *Firmicutes*, the other phylum with known PICIs, for the presence of these elements. Importantly, our analysis was able to identify cf-PICIs in different genera of *Firmicutes*, including *Lactococcus*, *Enterococcus*, *Bacillus*, *Clostridium*, and *Staphylococcus*. While the cf-PICIs in *Proteobacteria* showed a conserved genetic organization, with the morphogenetic genes (*capsid-protease-portal-head connectors*) first, followed by the ones involved in DNA recognition and packaging (*hnh-terS-terL*), two distinct groups were observed in the cf-PICIs from *Firmicutes*. One group was composed of the elements present in *Bacillus*, *Clostridium*, and *Staphylococcus*, while the second group was composed of the elements present in *Lactococcus* and *Enterococcus* (Figure S1; Table S2). While both groups had similar gene content, the organization/localization

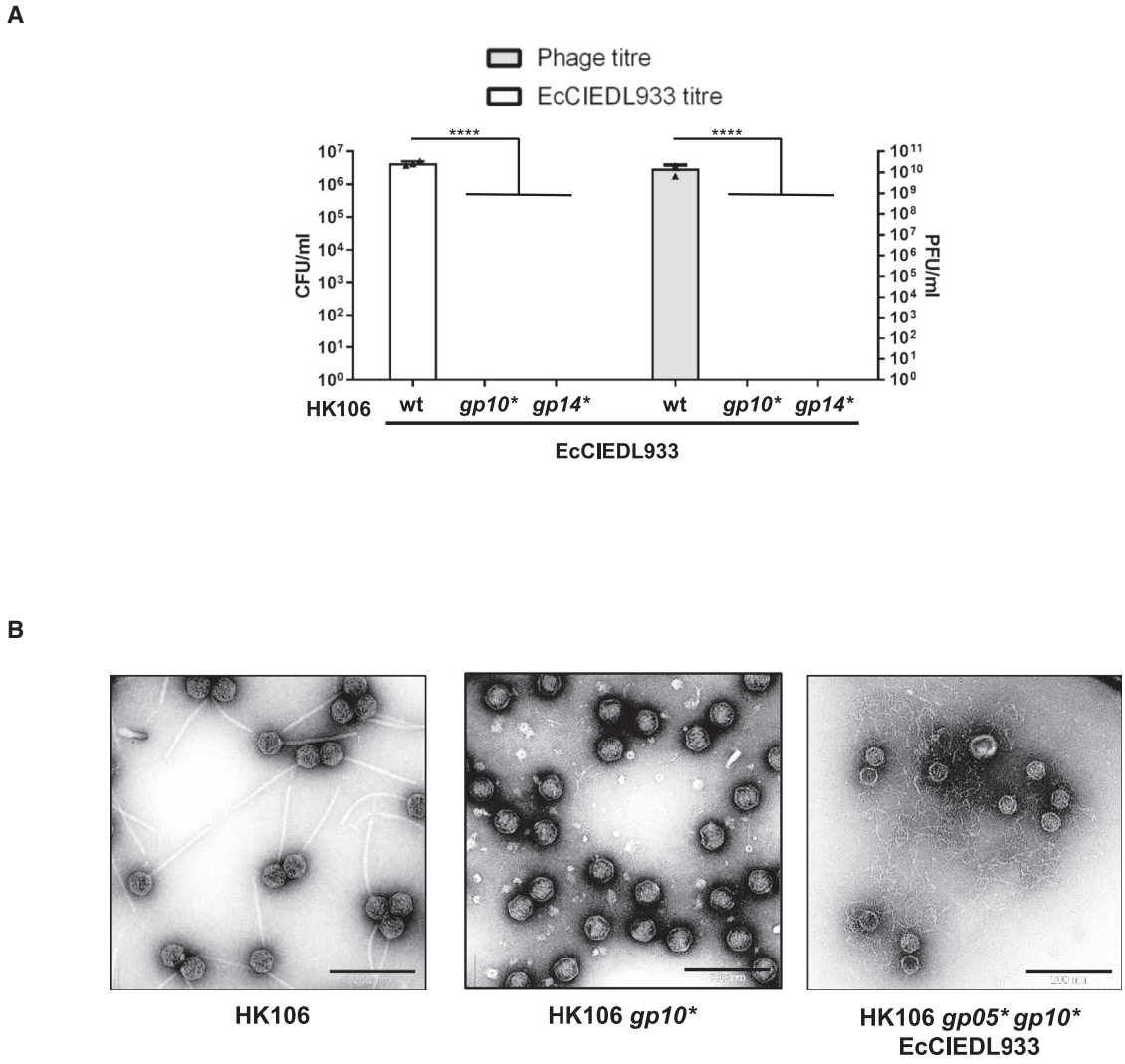


Figure 6. EcCIEDL933 requires phage tails to generate infective particles

(A) Lysogenic strains for phage HK106 (WT, *gp10** or *gp14**) carrying EcCIEDL933 were MC induced (2 μg/mL), and the phage and PICI titers were determined using *E. coli* 594 as the recipient strain. Note that *gp10* and *gp14* encode major tail and tape measure proteins, respectively, required for phage tail formation. The means of the colony-forming units (CFUs) or phage-forming units (PFUs) and SD of three independent experiments are presented (n = 3). The transduction and titration limit of detection is 10 CFUs or PFUs, respectively. A one-way ANOVA with Dunnett's multiple comparisons test was performed to compare mean differences between WT and individual mutants. Adjusted *p* values were as follows: ns > 0.05; *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001; ****p ≤ 0.0001.

(B) Electron microscopy analyses of different MC-induced HK106 lysogenic strains (WT, major tail mutant (*gp10**) or capsid (*gp05**) and major tail (*gp10**) mutants) in presence or absence of EcCIEDL933. Different fields are shown, containing only HK106 phage particles, HK106-sized capsid particles, or EcCIEDL933-sized capsid particles. Images were collected using a JEOL 1200 TEM microscope with 12K magnification. EcCIEDL933 particles have smaller heads. Scale bars represent 200 nm. See also Tables S4 and S5.

of these genes within the packaging operon was different. The genes encoding the portal, protease, and capsid proteins of the first group were present at the beginning of the operon, while the genes encoding HNH, TerS, and TerL start the operon of the second group. This suggests that cf-PICI may have different origins in these two phyla.

cf-PICIs arose three times independently

To detail the evolutionary history of cf-PICIs, we searched specifically for homologs of their capsid, HMP, and large terminase proteins in the database of complete phage genomes used in our initial analysis. In each case (capsid, HMP, or TerL), we collected the phage hits with highly significant e values (<$1e^{-10}$) (see STAR Methods). This allowed us to put together a list of homologs for the capsid (113 phages + 61 cf-PICIs), HMP (171 phages and 60 cf-PICIs), and TerL (420 phages + 60 cf-PICIs) proteins present in the cf-PICIs or in the phages. We then built the phylogenetic trees of the three types of proteins. The conclusions were identical for the analyses of the capsid and the terminase proteins and showed a few differences for the HMP (Figures 7, S6, and S7). We observed that, based on their capsid and TerL proteins, cf-PICIs were regrouped into three clades that are clearly

Figure 7. Capsid of cf-PICIs form three distinct phylogenetic groups and are evolutionarily separated from phage homologs

Phylogenetic trees inferred from the alignment of 61 capsid homologs from cf-PICIs and the best 113 phage capsid homologs. Phages branches are collapsed and shown as triangles. The different colors on the branches indicate the three different clades, formed by either *Gammaproteobacteria* (in blue), *Lactobacillus* (in

(legend continued on next page)

separated from each other by many phage genes. These clades have very high bootstrap support showing a clear separation from phages and the other cf-PICI clades (Figures 7 and S6). The HMP proteins showed a similar separation between three clades with excellent bootstrap support. However, it also showed a few scattered proteins for cf-PICI from *Firmicutes* that are intermingled with phage-encoded HMP (Figure S7). These branches tend to be poorly supported by the bootstrap analysis, and it is difficult at this stage to conclude if they represent gene flow between cf-PICI and phages or artifacts of phylogenetic inference. The three major clades supported by the three phylogenetic analyses correspond to the three groups distinguished in the previous section in terms of gene content and genetic organization, thereby indicating a concordance between the separation based on the phylogenetic analysis of protein sequences and the separation based on gene composition and order. Without excluding gene flow between cf-PICI and phages, these results strongly suggest an independent emergence of the three groups of cf-PICIs.

The analysis of each of the cf-PICI clade shows a clear separation between the satellites and the phage homologs, i.e., the clades only include cf-PICI components even if we deliberately selected the phage homologs that were most similar to the cf-PICI. Furthermore, the phages closest to each clade of cf-PICIs tend to be the same in the TerL and capsid tree. The consistent phylogeny for capsid and TerL components (and to a slightly lesser extent, HMPs) suggests strong genetic linkage between these functions. This further confirms that cf-PICIs are clearly distinguishable from the phages in the database and suggests a low frequency of genetic exchanges of these key components between cf-PICIs and phages.

Within each clade of cf-PICIs, one identifies elements present in the genomes of multiple bacterial genera. In the EcCIEDL933-containing clade, we could identify bacteria from very different *Gammaproteobacteria*, from *E. coli* to *Yersinia* or *Xenorhabdus*. Similarly, the clade containing the *S. aureus* element includes bacteria from *Bacillus* and *Clostridium*, which are very distantly related. This shows that either these elements have an exceptionally broad host range upon transfer through virions or that they are very ancient and have diversified within distinct bacterial taxonomic groups. The analysis of the distances in the phylogenetic trees shows very long branches between the cf-PICI and the closest phages even in the highly conserved protein TerL (Figure S6), which is also suggestive of an old origin for cf-PICIs. Hence, the cf-PICIs are neither recently derived from phages nor the result of recent prophage genetic degradation. Instead, they constitute deep lineages of elements that, compared with the other previously described PICIs, have more proteins that are homologous to those of phages.

DISCUSSION

Satellites are MGEs whose life cycle depends on a helper virus but lack extensive nucleotide sequence homology with them. So far, the origin of these elements or how have they evolved remains a mystery. In this manuscript, we report that a new family of phage satellites is able to specifically package their DNA into small-sized capsids, which are of satellite origin. In most of the previously analyzed satellites, the production of the small-sized capsids and the preferential packaging of the satellites are two independent processes. For example, some *cos* satellites (SaPIs and P4) produce small capsids, but they do not package their DNA preferentially since they carry the same *cos* sequence that is present in their helper phages.[9,15,16,30] Meanwhile, in the *pac* SaPIs, the phage- and the SaPI-encoded small terminases can interact with both the small and large capsids, a process that only limits phage reproduction when a partial phage genome is packaged into the small capsids.[14] Additionally, some *E. coli* PICIs that use large phage capsids for packaging express the Rpp protein, which by interacting with the phage-encoded TerS protein modifies its specificity, promoting the preferential package of the PICI DNA into large capsids.[24] Lastly, the recently described PLE satellites also produce small capsids, although the mechanism of small capsid production and how they preferentially package their genomes into these small-sized capsids is still unknown.[19]

Contrary to the previously described satellites, in the cf-PICI family of satellites, both the production of small capsids and the exclusive packaging of the elements in these small capsids are linked. In the system characterized here, phage and satellite packaging are clearly independent, and they have evolved to be also mutually exclusive (Figures 3 and 4). Why do different satellites use different strategies? Most of the previously characterized satellites hijack the machinery of helper phages that use the classical *cos* or *pac* systems for packaging. In these systems, the terminase complex, formed by the large and small subunits, interacts with their cognate *pac* or *cos* site to initiate packaging.[3,12] Interestingly, the new family of satellites described here hijacks a different family of phages, present both in *Proteobacteria* and *Firmicutes*, which require not only the TerS-TerL complex but also the activity of an additional player, the HNH protein, to initiate packaging (Figure 4).[15,31] A similar scenario was observed with SaPIbov5, which uses *S. aureus* phages as helpers encoding this new mechanism of packaging.[15] However, contrary to what is seen in SaPIbov5, which carries the same phage *cos* sequence for packaging and redirects capsid formation by the expression of the Ccm protein,[16,17] the members of the cf-PICI family have evolved to promote the formation of the small-sized capsids and their preferential packaging using a completely different strategy: instead of expressing a protein that either redirects helper phage capsid assembly or promotes PICI preferential packaging, these elements have incorporated in their genomes the phage genes required for capsid formation and packaging, which posteriorly have evolved to express proteins that do not cross talk with the phage-encoded ones. This would explain the exclusive packaging of these elements in the large (phage) or small (PICI) capsids.

Another interesting characteristic of the EcCIEDL933 element is that it does not interfere, at least in the conditions tested in these studies, with the reproduction of their helper phages.

orange), and a broader set of *Firmicutes* (in red). Bootstrap values for the main branches are indicated as gray circles. The tree was visualized and edited in iTOL.[33] Phylogenetic tree resulting from the alignment of capsid proteins from 61 cf-PICIs and 3,725 phages. Genetic organization of each cluster is shown. See also Figures S1, S6, and S7.

Since EcCIEDL933 is the only element of this new lineage characterized so far, we do not know if this behavior is exclusive of this satellite or is a characteristic of this new family. Since we have previously demonstrated that by interfering with phage reproduction PICIs drive phage evolution and ecology,[29] the lack of interference observed here opens new avenues by which phages and satellites may interact. In support of this idea, we have recently demonstrated that PICIs encoding anti-phage systems can be mobilized by some phages against other phages and MGEs, demonstrating that phage and PICIs can establish mutually beneficial relationships.[32] Since the function of most of the EcCIEDL933-encoded genes is unknown, we cannot disregard them as some of them could be involved in interfering with the reproduction of other phages. Indeed, other cf-PICIs encode abortive infection genes (Figure S1), suggesting that those cf-PICIs may play an important role in competition between MGEs. However, based on our results, one might speculate that encoding a specific packaging system may allow cf-PICIs to use the cell resources more efficiently than the classical PICIs, thus alleviating the cost of their production on the replicating helper phage. Additionally, cf-PICIs only require phage tails to complete the PICI infective particle. If we assume that tails are produced in excess, this strategy mitigates any impact that cf-PICIs could have on the phage. Indeed, other satellites require the production of proteins to change capsid sizes, and some of them cannot preclude the (non-productive) partial packaging of phages in the small capsids. Thus, cf-PICIs may ensure an optimal use of resources to produce their PICI particles and have a negligible impact on the production of viral particles. Independent of the cost they produce in their helper phages, it is interesting to remark that EcCIEDL933 and most *E. coli* cf-PICIs encode genes involved in virulence (*rdg*R and/or *per*C). Therefore, in addition to their impact on phage ecology, cf-PICIs can play other relevant roles in bacterial virulence.

Importantly, this new strategy of parasitism is present in satellites from *Proteobacteria* and *Firmicutes*, confirming its significance in nature. Their phylogeny and the comparative analyses of the gene repertoires show that they are clearly distinct from phages, even for the proteins that have homologs in both types of elements. The phylogenetic analysis further shows that there are at least three groups of cf-PICIs that emerged independently raising interesting questions in relation to their origin. One alternative is that degenerated phages led to the independent emergence of these satellites. However, several results argue against this hypothesis. First, there is very low homology between cf-PICI and known phages for most genes. Second, the homology is restricted to the integrase-distal part of the element. Third, the least homologous region shows conserved genetic organization across *Firmicutes* cf-PICIs, even in those of different clades. Finally, the integrase-proximal organization of cf-PICI shares similarities with those of classical PICIs. A comparative analysis of PICIs and cf-PICIs is beyond the scope of this study, but very recent results suggest that these elements have a common origin and share some genes by recombination while clearly differing from phages.[7] Hence, we hypothesize that the capsid and packaging genetic modules were likely obtained from different phages and incorporated together into ancestral PICIs present in different bacterial species. Given the evolutionary distances between cf-PICI and phages, even in the highly conserved large subunit of the terminase, the acquisition of the phage capsid and packaging modules must have occurred a long time ago. Multiple independent acquisitions of the modules from different phages explain the phylogenetic patterns and our observation of different genetic organizations in different groups. These modules then evolved to adapt to the cf-PICI biology, which may have resulted in the elimination of accessory genes to fit the small cf-PICI genomes, leading to a minimal capsid module that has remarkably similar types of components in the three groups. These components may correspond to a minimal set of genes necessary to build a capsid.

Once again, PICIs have provided unexpected insights into the biology of the satellite viruses. Because of the simplicity of the experimental models, evolutionary studies with phages and their satellites can give insights into the general principles of the biology of the eukaryotic viruses and their satellites. We have applied this concept here and have demonstrated the existence of a novel way by which satellites and helper phages interact. We anticipate that a better characterization of this interaction will contribute to our understanding on how viruses and satellites evolve and how this evolution and coexistence impacts on the biology of both the prokaryotic and eukaryotic cells infected by these elements.

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
  - Bacterial strains and growth conditions
- METHOD DETAILS
  - Plasmid construction
  - DNA methods
  - Phage plaque assay
  - Phage and PICI induction
  - Phage titration
  - PICI transduction
  - Southern Blot
  - Phage and PICI particles precipitation
  - Phage and PICI DNA extraction
  - Whole genome sequencing (phage and/or PICI DNA from capsid extraction)
  - Electron microscopy
  - EcCIEDL933 protein identification
  - Identification of PICIs and KEGG analysis
  - Phylogenetic analyses
  - Weighted Gene Repertoire Relatedness
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chom.2022.12.001.

ACKNOWLEDGMENTS

This work was supported by grants MR/M003876/1, MR/V000772/1, and MR/S00940X/1 from the Medical Research Council (UK); BB/N002873/1, BB/V002376/1, BB/V009583/1, and BB/S003835/1 from the Biotechnology and Biological Sciences Research Council (BBSRC, UK); ERC-ADG-2014 Proposal no. 670932 Dut-signal (from EU); and Wellcome Trust 201531/Z/16/Z to J.R.P.; and grants INCEPTION project (PIA/ANR-16-CONV-0005); Equipe FRM (Fondation pour la Recherche Médicale): EQU201903007835; Laboratoire d'Excellence IBEIDIntegrative Biology of Emerging Infectious Diseases (ANR-10-LABX-62-IBEID); and SALMOPROPHAGEANR-16-CE16-0029 to E.P.C.R. L.M.-R. was the recipient of a Spanish postdoctoral fellowship from the Fundación Ramón Areces (2018–2020). A.F.-S. is the recipient of a fellowship from Imperial College.

AUTHOR CONTRIBUTIONS

J.R.P. conceived the study; N.A., L.M.-R., and A.F.-S. conducted the experiments; J.M.d.S. and E.P.C.R. performed the genomic analyses; N.A., L.M.-R., J.M.d.S., J.C., E.P.C.R., A.F.-S., and J.R.P. analyzed the data. E.P.C.R., A.F.-S., and J.R.P. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 19, 2022
Revised: October 6, 2022
Accepted: December 1, 2022
Published: January 2, 2023

REFERENCES

1. Koonin, E.V., Makarova, K.S., Wolf, Y.I., and Krupovic, M. (2020). Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. Nat. Rev. Genet. *21*, 119–131. https://doi.org/10.1038/s41576-019-0172-9.
2. Moura de Sousa, J.A., and Rocha, E.P.C. (2020). To catch a hijacker: abundance, evolution and genetic diversity of P4-like bacteriophage satellites. Philos. Trans. R. Soc. B Biol. Sci. *377*. 20200475. https://doi.org/10.1098/rstb.2020.0475.
3. Fillol-Salom, A., Martínez-Rubio, R., Abdulrahman, R.F., Chen, J., Davies, R., and Penadés, J.R. (2018). Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. ISME J. *12*, 2114–2128. https://doi.org/10.1038/s41396-018-0156-3.
4. Martínez-Rubio, R., Quiles-Puchalt, N., Martí, M., Humphrey, S., Ram, G., Smyth, D., Chen, J., Novick, R.P., and Penadés, J.R. (2017). Phage-inducible islands in the Gram-positive cocci. ISME J. *11*, 1029–1042. https://doi.org/10.1038/ismej.2016.163.
5. O'Hara, B.J., Barth, Z.K., McKitterick, A.C., and Seed, K.D. (2017). A highly specific phage defense system is a conserved feature of the *Vibrio cholerae* mobilome. PLoS Genet. *13*. e1006838. https://doi.org/10.1371/journal.pgen.1006838.
6. Barth, Z.K., Silvas, T.V., Angermeyer, A., and Seed, K.D. (2020). Genome replication dynamics of a bacteriophage and its satellite reveal strategies for parasitism and viral restriction. Nucleic Acids Res. *48*, 249–263. https://doi.org/10.1093/nar/gkz1005.
7. Moura de Sousa, J.A., Fillol-Salom, A., Penadés, J.R., and Rocha, E.P.C. (2022). Identification and characterization of thousands of bacteriophage satellites across bacteria. https://doi.org/10.1101/2022.09.14.508007.
8. Penadés, J.R., and Christie, G.E. (2015). The phage-inducible chromosomal islands: a family of highly evolved molecular parasites. Annu. Rev. Virol. *2*, 181–201. https://doi.org/10.1146/annurev-virology-031413-085446.
9. Christie, G.E., and Dokland, T. (2012). Pirates of the Caudovirales. Virology *434*, 210–221. https://doi.org/10.1016/j.virol.2012.10.028.
10. Ibarra-Chávez, R., Hansen, M.F., Pinilla-Redondo, R., Seed, K.D., and Trivedi, U. (2021). Phage satellites and their emerging applications in biotechnology. FEMS Microbiol. Rev. *45*, fuab031. https://doi.org/10.1093/femsre/fuab031.
11. Agarwal, M., Arthur, M., Arbeit, R.D., and Goldstein, R. (1990). Regulation of icosahedral virion capsid size by the in vivo activity of a cloned gene product. Proc. Natl. Acad. Sci. USA *87*, 2428–2432. https://doi.org/10.1073/pnas.87.7.2428.
12. Fillol-Salom, A., Miguel-Romero, L., Marina, A., Chen, J., and Penadés, J.R. (2020). Beyond the CRISPR-Cas safeguard: PICI-encoded innate immune systems protect bacteria from bacteriophage predation. Curr. Opin. Microbiol. *56*, 52–58. https://doi.org/10.1016/j.mib.2020.06.002.
13. Dearborn, A.D., Wall, E.A., Kizziah, J.L., Klenow, L., Parker, L.K., Manning, K.A., Spilman, M.S., Spear, J.M., Christie, G.E., and Dokland, T. (2017). Competing scaffolding proteins determine capsid size during mobilization of *Staphylococcus aureus* pathogenicity islands. eLife, e30822. https://doi.org/10.7554/elife.30822.
14. Ubeda, C., Maiques, E., Tormo, M.A., Campoy, S., Lasa, I., Barbé, J., Novick, R.P., and Penadés, J.R. (2007). SaPI operon I is required for SaPI packaging and is controlled by LexA. Mol. Microbiol. *65*, 41–50. https://doi.org/10.1111/j.1365-2958.2007.05758.x.
15. Quiles-Puchalt, N., Carpena, N., Alonso, J.C., Novick, R.P., Marina, A., and Penadés, J.R. (2014). Staphylococcal pathogenicity island DNA packaging system involving *cos*-site packaging and phage-encoded HNH endonucleases. Proc. Natl. Acad. Sci. USA *111*, 6016–6021. https://doi.org/10.1073/pnas.1320538111.
16. Carpena, N., Manning, K.A., Dokland, T., Marina, A., and Penadés, J.R. (2016). Convergent evolution of pathogenicity islands in helper cos phage interference. Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*. 20150505. https://doi.org/10.1098/rstb.2015.0505.
17. Hawkins, N.C., Kizziah, J.L., Penadés, J.R., and Dokland, T. (2021). Shape shifter: redirection of prolate phage capsid assembly by staphylococcal pathogenicity islands. Nat. Commun. *12*, 6408. https://doi.org/10.1038/s41467-021-26759-x.
18. Matos, R.C., Lapaque, N., Rigottier-Gois, L., Debarbieux, L., Meylheuc, T., Gonzalez-Zorn, B., Repoila, F., Lopes, Mde F., and Serror, P. (2013). *Enterococcus faecalis* prophage dynamics and contributions to pathogenic traits. PLoS Genet. *9*. e1003539. https://doi.org/10.1371/journal.pgen.1003539.
19. Netter, Z., Boyd, C.M., Silvas, T.V., and Seed, K.D. (2021). A phage satellite tunes inducing phage gene expression using a domesticated endonuclease to balance inhibition and virion hijacking. Nucleic Acids Res. *49*, 4386–4401. https://doi.org/10.1093/nar/gkab207.
20. Ziermann, R., and Calendar, R. (1990). Characterization of the cos sites of bacteriophages P2 and P4. Gene *96*, 9–15. https://doi.org/10.1016/0378-1119(90)90334-n.
21. Chen, J., Ram, G., Penadés, J.R., Brown, S., and Novick, R.P. (2015). Pathogenicity island-directed transfer of unlinked chromosomal virulence genes. Mol. Cell *57*, 138–149. https://doi.org/10.1016/j.molcel.2014.11.011.
22. Ram, G., Chen, J., Ross, H.F., and Novick, R.P. (2014). Precisely modulated pathogenicity island interference with late phage gene transcription. Proc. Natl. Acad. Sci. USA *111*, 14536–14541. https://doi.org/10.1073/pnas.1406749111.
23. Ram, G., Chen, J., Kumar, K., Ross, H.F., Ubeda, C., Damle, P.K., Lane, K.D., Penadés, J.R., Christie, G.E., and Novick, R.P. (2012). Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. Proc. Natl. Acad. Sci. USA *109*, 16300–16305. https://doi.org/10.1073/pnas.1204615109.
24. Fillol-Salom, A., Bacarizo, J., Alqasmi, M., Ciges-Tomas, J.R., Martínez-Rubio, R., Roszak, A.W., Cogdell, R.J., Chen, J., Marina, A., and Penadés, J.R. (2019). Hijacking the hijackers: *Escherichia coli* pathogenicity islands redirect helper phage packaging for their own benefit. Mol. Cell *75*. 1020–1030.e4. https://doi.org/10.1016/j.molcel.2019.06.017.

25. Moura de Sousa, J.A., Pfeifer, E., Touchon, M., and Rocha, E.P.C. (2021). Causes and consequences of bacteriophage diversification via genetic exchanges across lifestyles and bacterial taxa. Mol. Biol. Evol. *38*, 2497–2512. https://doi.org/10.1093/molbev/msab044.

26. Perna, N.T., Plunkett, G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., et al. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature *409*, 529–533. https://doi.org/10.1038/35054089.

27. Humphrey, S., Fillol-Salom, A., Quiles-Puchalt, N., Ibarra-Chávez, R., Haag, A.F., Chen, J., and Penadés, J.R. (2021). Bacterial chromosomal mobility via lateral transduction exceeds that of classical mobile genetic elements. Nat. Commun. *12*, 6509. https://doi.org/10.1038/s41467-021-26004-5.

28. Asadulghani, M., Ogura, Y., Ooka, T., Itoh, T., Sawaguchi, A., Iguchi, A., Nakayama, K., and Hayashi, T. (2009). The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. PLoS Pathog. *5*. e1000408. https://doi.org/10.1371/journal.ppat.1000408.

29. Frígols, B., Quiles-Puchalt, N., Mir-Sanchis, I., Donderis, J., Elena, S.F., Buckling, A., Novick, R.P., Marina, A., and Penadés, J.R. (2015). Virus satellites drive viral evolution and ecology. PLoS Genet. *11*. e1005609. https://doi.org/10.1371/journal.pgen.1005609.

30. Lindqvist, B.H., Dehò, G., and Calendar, R. (1993). Mechanisms of genome propagation and helper exploitation by satellite phage P4. Microbiol. Rev. *57*, 683–702. https://doi.org/10.1128/mr.57.3.683-702.1993.

31. Kala, S., Cumby, N., Sadowski, P.D., Hyder, B.Z., Kanelis, V., Davidson, A.R., and Maxwell, K.L. (2014). HNH proteins are a widespread component of phage DNA packaging machines. Proc. Natl. Acad. Sci. USA *111*, 6022–6027. https://doi.org/10.1073/pnas.1320952111.

32. Fillol-Salom, A., Rostøl, J.T., Ojiogu, A.D., Chen, J., Douce, G., Humphrey, S., and Penadés, J.R. (2022). Bacteriophages benefit from mobilizing pathogenicity islands encoding immune systems against competitors. Cell *185*. 3248–3262.e20. https://doi.org/10.1016/j.cell.2022.07.014.

33. Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296. https://doi.org/10.1093/nar/gkab301.

34. Hoffmann, S., Schmidt, C., Walter, S., Bender, J.K., and Gerlach, R.G. (2017). Scarless deletion of up to seven methyl-accepting chemotaxis genes with an optimized method highlights key function of CheM in *Salmonella* Typhimurium. PLoS One *12*. e0172630. https://doi.org/10.1371/journal.pone.0172630.

35. Blank, K., Hensel, M., and Gerlach, R.G. (2011). Rapid and highly efficient method for scarless mutagenesis within the *Salmonella enterica* chromosome. PLoS One *6*, e15763. https://doi.org/10.1371/journal.pone.0015763.

36. Solano, C., García, B., Latasa, C., Toledo-Arana, A., Zorraquino, V., Valle, J., Casals, J., Pedroso, E., and Lasa, I. (2009). Genetic reductionist approach for dissecting individual roles of GGDEF proteins within the c-di-GMP signaling network in *Salmonella*. Proc. Natl. Acad. Sci. USA *106*, 7997–8002. https://doi.org/10.1073/pnas.0812573106.

37. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60. https://doi.org/10.1038/nmeth.3176.

38. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

39. Steenwyk, J.L., Buida, T.J., Li, Y., Shen, X.-X., and Rokas, A. (2020). ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. *18*. e3001007. https://doi.org/10.1371/journal.pbio.3001007.

40. Nguyen, L.-T., Schmidt, H.A., Haeseler von, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274. https://doi.org/10.1093/molbev/msu300.

41. Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. *35*, 1026–1028. https://doi.org/10.1038/nbt.3988.

42. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

43. Perkins, D.N., Pappin, D.J.C., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

44. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. *34*, D354–D357. https://doi.org/10.1093/nar/gkj102.

45. Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011). Easyfig: a genome comparison visualizer. Bioinformatics *27*, 1009–1010. https://doi.org/10.1093/bioinformatics/btr039.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Digoxigenin-AP Fab fragments	Roche	11093274910
Bacterial and virus strains		
Bacteriophages, see Table S3	N/A	N/A
Bacterial strains, see Table S3	N/A	N/A
Chemicals, peptides, and recombinant proteins		
Bacteriological agar	VWR Chemicals	84609.05; CAS 9002-18-0
Nutrient Broth No. 2	Thermo Fisher Scientific (Thermo Scientific™)	10259632
Platinum® Taq DNA Polymerase High Fidelity	Thermo Fisher Scientific (Invitrogen™)	Cat#11304011
DreamTaq DNA Polymerase	Thermo Fisher Scientific (Thermo Scientific™)	Cat#EP0703
T4 DNA Ligase	Thermo Fisher Scientific (Invitrogen™)	10786591
Ampicillin sodium salt	Merk (Sigma-Aldrich)	A9518; CAS 69-52-3
Chloramphenicol	Merk (Sigma-Aldrich)	C0378; CAS 56-75-7
Kanamycin Sulfate	Merk (Sigma-Aldrich)	60615; CAS 70560-51-9
Mitomycin C	Merk (Sigma-Aldrich)	M0503; CAS 50-07-7
L-(+)-Arabinose	Merk (Sigma-Aldrich)	A3256; CAS 5328-37-0
Thermo Scientific X-Gal	Thermo Fisher Scientific (Thermo Scientific™)	10490470
Anhydrotetracycline hydrochloride	Merk (VETRANAL®)	37919; CAS 13803-65-1
InstantBlue Ultrafast Protein Stain	Sigma	Cat#ISB1L-1L
Pierce Unstained Protein MW Marker	Thermo Fisher Scientific (Thermo Scientific™)	Cat#26610
Cesium chloride	Sigma	Cat#C4036-100G
Poly (ethylene glycol) 8000	Sigma	Cat#89510-1KG-F
Uranyl formate 99.9-100%	VWR	22450, CAS 16984-59-1
Paraformaldehyde powder	Fisher Scientific	Cat#10131580
30% Acrylamide/Bis Solution, 37.5:1	Biorad	Cat#1610158
Ammonium persulfate	Biorad	Cat#1610700
N, N, N', N'-Tetramethylethylenediamine	Sigma	Cat#T9281-25ml
Phenol:Chloroform:Isoamyl alcohol 25:24:1	Sigma	Cat#P2069-100ML
Digoxigenin-11-dUTP, alkali-stable	Roche	11093088910
Nylon Membranes, positively charged	Roche	11417240001
Critical commercial assays		
QIAquick PCR Purification Kit	QIAgen	Cat#28106
QIAprep Spin Miniprep Kit	QIAgen	Cat#27106
Gel Elute Bacterial genomic DNA kit	Sigma	Cat#NA2121-1KT
Mix2Seq Kit NightXpress	Eurofins Genomics	N/A
Deposited data		
Illumina Whole Genome Sequencing	This paper	http://www.ncbi.nlm.nih.gov/bioproject/884654
Original data in Mendeley dataset	This paper	https://doi.org/10.17632/7r88m4fp6c.1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
Primers used in this study, see Table S7	This paper	N/A
Recombinant DNA		
Plasmids used in this study, see Table S7	This paper	N/A
Software and algorithms		
Image J	Schneider et al.[42]	https://imagej.nih.gov/ij/
Mascot	Perkins et al.[43]	www.matrixscience.com
GraphPad prism	GraphPad Software 9.3. 1	https://www.graphpad.com/scientific-software/prism/
Clustal Omega	Multiple Sequence Alignment	https://www.ebi.ac.uk/Tools/msa/clustalo/
KEGG	Kanehisa et al.[44]	http://www.genome.jp/kegg
mmseqs2	Steinegger and Söding[41]	13-45111
Diamond	Buchfink et al.[37]	v2.0.6.144
mafft-linsi	Katoh and Standley[38]	v7.490
Clipkit	Steenwyk et al.[39]	v1.3.0
Iqtree	Nguyen et al.[40]	v1.6.12
iTol	Letunic and Bork[33]	v6.5.2
EasyFig	Sullivan et al.[45]	http://mjsull.github.io/Easyfig/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, José R Penadés (j.penades@imperial.ac.uk).

Materials availability

Strains, phages and plasmids generated in this study are available upon request and without restrictions from the lead contact upon request.

Data and code availability

- Illumina Whole Genome Sequencing data have been deposited at the NCBI SRA database under accession code BioProject PRJNA884622 and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Original agarose gel, Southern blot, electron microscopy, SDS PAGE gel and phage spot assay images have been deposited at Mendeley and are publicly available as of the date of publication. The DOI is listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial strains and growth conditions

Phages and bacterial strains used in this study are listed in Table S3. *E. coli* strains were grown at 37°C or 30°C on Luria-Bertani (LB) agar or in LB broth with shaking (120 r.p.m.) supplemented with Ampicillin (100 μg ml^{-1}), Kanamycin (30 μg ml^{-1}) or Chloramphenicol (20 μg ml^{-1}; all Sigma-Aldrich) when required.

METHOD DETAILS

Plasmid construction

The plasmids used in this study (Table S7) were constructed by cloning PCR products, amplified with primers listed in Table S7, into the pBAD18 vector using enzymatic digestion and ligation. Plasmids were verified by Sanger sequencing in Eurofins Genomics.

DNA methods

The introduction of the chloramphenicol (*cat*) resistance marker into the EcCIEDL933 element was performed as described[33] using λ Red recombinase-mediated recombination. Briefly, the marker was amplified by PCR using primers listed in Table S7 from plasmid pKD3, and the PCR product was transformed into the recipient strain harbouring plasmid pRWG99, which expresses the λ Red recombinase, and the marker was inserted into the PICI genome. The insertion of the resistance marker, which replaces the 1795-1976 genes (see Figure 1A), was verified by PCR.

For PICI mutagenesis, the site-directed scarless mutagenesis was performed as described previously.[34,35] Briefly, the *km*R marker together with an I-*Sce*I recognition restriction site was amplified by PCR, using primers listed in Table S7. Then, the PCR product was inserted into the recipient strain harbouring plasmid pRWG99, which expresses the λ Red recombinase protein. After verification of the insertion by PCR, 80mer DNA fragments derived from oligonucleotides which contains the desired mutation were electroporated into the mutant strain expressing the λ Red recombinase-mediated system, and successful recombinants were selected by expression of I-*Sce*I endonuclease. The different PICI mutants obtained were verified by PCR and Sanger sequencing in Eurofins Genomics.

For phage mutagenesis, an allelic replacement strategy was performed.[36] Briefly, the allelic-exchange vector, pK03-Blue, was used to clone the desired genes using the primers listed in Table S7. Then, the plasmid was introduced into the strain carrying the prophage and the transformants were selected on LBA plates supplemented with chloramphenicol and incubated at 32°C for the selection of the temperature-sensitive plasmid. To produce the homologous recombination, the plasmid was forced to integrate into the phage genome at the non-permissive temperature (42°C). Light blue colonies, which are indicative of plasmid integration, were grown in LB broth at 32°C and ten-fold serial dilution of the overnight cultures was plated on LBA plated containing X-gal (5-bromo-4-chloro- 3-indolyl-B-D-galactopyranoside) and sucrose 5% (to force the plasmid loss) and incubated at 32°C for 24 h. White colonies, which is indicative that the plasmid is loss, were screened for chloramphenicol sensitivity. The different phage mutants obtained were verified by PCR and Sanger sequencing in Eurofins Genomics.

Phage plaque assay

A 1:50 dilution (in fresh LB broth) of the overnight strains were grown until an OD_{600}=0.34 was reached. Bacterial lawns were prepared by mixing 300 μL of cells with phage top agar (PTA) and poured onto square plates. Then, serial dilutions of phages were prepared in phage buffer (50mM Tris pH 8, 1mM $MgSO_4$, 4mM $CaCl_2$ and 100mM NaCl) and spotted on correspond plate, which were incubated at 37°C for 24h.

Phage and PICI induction

For PICI and phage induction, overnight cultures of lysogenic donor strains in presence and absence of the PICI were diluted 1:50 in fresh LB broth and grown at 37°C and 120 r.p.m. until an OD_{600} of 0.2 was reached. Mitomycin C (2 μg ml^{-1}; Sigma-Aldrich from *Streptomyces caespitosus*) was added to induce the prophage and the induced cultures were grown at 32°C with slow shaking (80 r.p.m.). Generally, cell lysis occurred 4-5 h post-induction and the induced samples were filtered using sterile 0.2 μm filters (Minisart® single use syringe filter unit, hydrophilic and non-pyrogenic, Sartonium Stedim Biotech). The number of phage or PICI particles in the resultant lysate was quantified.

Phage titration

The number of phage particles were quantified using the tittering assay. Briefly, a 1:50 dilution (in fresh LB broth) of an overnight recipient strain was grown until an OD_{600} of 0.34 was reached. Then, recipient strain was infected using 100 μL of cells with the addition of 100 μL of phage lysate serial dilutions prepared with phage buffer (50mM Tris pH 8, 1mM $MgSO_4$, 4mM $CaCl_2$ and 100mM NaCl), and incubated for 5 min at room temperature. The different mixtures of culture-phage were plated out on phage base agar plates (PBA; 25 g of Nutrient Broth No. 2, Oxoid; 7g agar) supplemented with $CaCl_2$ to a final concentration of 10mM. Plates were incubated at 37°C for 24h and the number of plaques formed (phage particles present in the lysate) were counted and represented as the plaque forming units (PFU/mL). The titration limit of detection is 10 PFUs.

PICI transduction

The number of PICI particles were quantified using the transduction tittering assay. Briefly, a 1:50 dilution (in fresh LB broth) of an overnight recipient strain was grown until an OD_{600} of 1.4 was reached. Then, strains were infected using 1 mL of recipient cells with the addition of 100 μL of PICI lysate serial dilutions prepared with phage buffer (50mM Tris pH 8, 1mM $MgSO_4$, 4mM $CaCl_2$ and 100mM NaCl), and cultures were supplemented with $CaCl_2$ to a final concentration of 4.4 mM before incubation for 30 min at 37°C. The different mixtures of culture-PICI were plated out on LBA plates containing the appropriate antibiotic. LBA plates were incubated at 37°C for 24h and the number of colonies formed (transduction particles present in the lysate) were counted and represented as the colony forming units (CFU/mL). The transduction limit of detection is 10 CFUs.

The *E. coli* 594 EcCIEDL933 *1795-1796*::*cat* was constructed by mobilising the EcCIEDL933 *cat* from the EDL933 original strain. Then, PCRs were performed to confirm the presence of EcCIEDL933 and that none of the resident prophages in EDL933 were transduced to the 594 strain.

Southern Blot

Following phage and PICI induction, one millilitre of each sample was taken at defined time points and pelleted. Samples were frozen at -20°C until all collections were accomplished. Then, samples were re-suspended in 50 μl lysis buffer (47.5 μl TES-Sucrose and 2.5 μl lysozyme [10 μg ml^{-1}]; Sigma-Aldrich) and incubated at 37°C for 1h. Then, 55 μl of SDS 2% proteinase K buffer (47.25 μl H_2O, 5.25 μl SDS 20%, 2.5 μl proteinase K [20 mg ml^{-1}], Sigma-Aldrich from *Tritirachium album*) was added to the lysates and incubated at 55°C for 30 min. Then, samples were vortexed with 10 μl of 10x loading dye for 1h. Following this incubation, samples were frozen and thawed in cycles of 5 min incubation in dry ice with ethanol and in a water bath at 65°C. This was repeated three times. To separate the chromosomal DNA, samples were run on 0.7% agarose gel at 30V, overnight. Then, the DNA was transferred to Nylon membranes (Hybond-N 0.45 mm pore size filters; Amersham Life Science) using standard methods. DNA was detected using a DIG-labelled probe (Digoxigenin-11-dUTP alkali-labile; Roche) and anti-DIG antibody (Anti-Digoxigenin-AP Fab fragments; Roche), before washing and visualisation. The primers used to obtain the DIG-labelled probes are listed in Table S7.

Phage and PICI particles precipitation

A large-scale induction was performed to precipitate phage and PICI particles and obtain their dsDNA. Briefly, a total of 100 ml lysate was produced by MC induction. Then, lysates were filtered using sterile 0.2 μm filters and treated with RNase (1 μg ml^{-1}) and DNase (1 μg ml^{-1}) for 30 min at room temperature. Afterwards, 1 M of NaCl was added to the lysate and incubated 1 h on ice. After incubation, the mix was centrifuged at 11,000 x *g* for 10 min at 4°C, and phages/PICIs were mixed with 10% wt/vol polyethylene glycol (PEG) 8000 and kept overnight at 4°C. Then, phages/PICIs were collected at 11,000 x *g* for 10 min at 4°C and the pellet was resuspended in 1 ml of phage buffer.

Phage and PICI DNA extraction

To extract the dsDNA, lysate from phage and PICI precipitation was treated with DNase (10 μg ml^{-1}) for 30 min at room temperature. Then, lysate was combined with an equal volume of lysis mix (2% SDS and 90 μg ml^{-1} proteinase K) and incubated at 55°C for 1 h. DNA was extracted with an equal volume of phenol:chloroform:isoamyl alcohol 25:24:1 and samples were centrifuged at 12,000 x *g* for 5 min, and the aqueous phase containing the DNA was obtained. The DNA was precipitated by 0.3M NaOAc and 2.25 volume of 100% ethanol, then pelleted at 12,000 x g for 30 min at 4°C and washed once with 1 ml of 70% ethanol. After centrifugation, the DNA pellets were air dried for 30 min and resuspended in 100 μl nuclease free water.

Whole genome sequencing (phage and/or PICI DNA from capsid extraction)

A total of 200 ml phage and/or PICI lysate was produced by MC induction. Particles were precipitates as previously explained. Then, the precipitated phages and/or PICIs were loaded on the CsCl step gradients (1.35, 1.5 and 1.7 g ml^{-1} fractions) and centrifuged at 80,000 x *g* for 2h at 4°C. The phage and/or PICI band was extracted from the CsCl gradients using a 23-gauge needle and syringe. Phages were dialyzed overnight to remove CsCl excess using SnakeSkin™ Dialysis Tubing (3.5K MWCO, 16mm dry) into 50mM of Tris pH 8 and 150mM NaCl buffer.

Following CsCl purification, phage and/or PICI lysates were treated with DNase (10 μg ml^{-1}) for 30 min at room temperature and DNase was inactivated with 5mM EDTA for 10 min at 70°C. Phage and/or PICI lysate was combined with an equal volume of lysis mix (2% SDS and 90 μg ml^{-1} proteinase K) and incubated at 55°C for 1 h. DNA was extracted with an equal volume of phenol:chloroform:isoamyl alcohol 25:24:1 and samples were centrifuged at 12,000 x *g* for 5 min, and the aqueous phase containing the DNA was obtained. DNA was ethanol precipitated as already described, and the resulting DNA was resuspended in 50 μl nuclease free water.

WGS was performed in SeqCenter company (USA) using Illumina NextSeq 2000 platform, obtaining 2x151bp pair end reads. A minimum size of 200 Mbp sequencing package was obtained. Trimmed reads were mapped to the appropriate genome containing the phage HK106 and EcCIEDL933. Only one replicate was sequenced per experiment.

Electron microscopy

To analyse the lysate by transmission electron microscopy (TEM), lysates were CsCl purified as described previously. Ten microliters of the dialysed samples were incubated in a carbon-coated gold grid for 5 min. Then, samples were fixed with 1% Paraformaldehyde for 2 min, before washing three times with distilled water for 30 sec. Afterwards, the samples were stained with 2% Uranyl Formate for 30 sec, and they were allowed to dry at room temperature for 15 min. The JEOL 1200 TEM microscope was used to examine the samples in the electron microscopy facility in School of Life Sciences MVLS of University of Glasgow. Photos were obtained at 12K of magnification and the particles measurement was done using the ImageJ software and Excel to obtain the mean and standard deviation.

EcCIEDL933 protein identification

Cf-PICI particles were obtained after particles precipitation and CsCl purification. Then, they were loaded in a 12% Acrylamide/Bis gel and analysed by SDS PAGE. The bands corresponding with single proteins were visualised using a Coomassie-based stain (InstantBlue Ultrafast Protein Stain), cut and stored in sterile water for protein identification. Mass Spectrometry for protein identification with the gel bands and the whole EcCIEDL933 particles were performed in BSRC Mass Spectrometry and Proteomics Facility in University of St. Andrews, UK. The gel bands were excised and cut into 1 mm cubes, then were subjected to in-gel digestion using a CEM in-gel digestion robot. Briefly, the gel cubes were faded by washing them with acetonitrile and digested with trypsin at 37°C. The whole EcCIEDL933 particles sample was also digested with trypsin. After the trypsin digestion, the peptides were extracted with 1%

formic acid and concentrated down to 20 μL using SpeedVac. They were then separated using a Thermo Ultimate nanoLC equipped with PepMap C18 trap & Easyspray 15 cm column, using a 60 min elution profile, with a gradient of increasing acetonitrile, containing 0.1 % formic acid to elute the peptides. The eluent was sprayed into Thermo Fusion Lumos mass spectrometer and analysed in Data Dependent Acquisition (DDA) mode performing 100 msec of MS followed by 80 msec Ms/MS analyses of the 20 most intense peaks seen in MS. MS/MS data for doubly and triply charged precursor ions was converted to using proteowizard msconvert and the data file generated was analysed using Mascot 6.2 search against a custom database BMS 220804 (6,517 sequences) which included the EcCIEDL933 and HK106 proteins. A tolerance 20ppm and 0.6 Da fragment ions was used.

Identification of PICIs and KEGG analysis

The KEGG analysis performed previously[3] suggested that the EcCIEDL933 element is the prototypical member of a new PICI family. To identify more members of this family, we performed a BLASTN, using EcCIEDL933 as a template and later other cf-PICIs, versus the NCBI database (with standard parameters, E value < 0.05). The positive hits with a conserved 3' end region (the packaging region) but a diverse 5' end region (integration, regulation, and replication region) were manually checked, to confirm that they were PICIs and no other elements such as phages. Also, we performed a BLASTN versus the NCBI database (with standard parameters, E value < 0.05) using the *int*-alpA region from the classical PICIs and looking for elements carrying a similar packaging module than that present in EcCIEDL933. To identify new members of this family in other species (*S. aureus* or *Lactobacillus*), we searched on the NCBI database for elements with a size comprised between 10-15 kb, encoding a divergent pair or transcriptional regulators next to an *int* gene,[4] and a packaging module similar to that present in EcCIEDL933. Additionally, these elements should have unique attachment sites that are never occupied by prophages, and they should lack phage lytic and tail genes. This conservative search was followed by the analysis of orthologues performed for the EcCIEDL933 element (Table S1), which corroborates that the newly identified elements correspond to cf-PICIs. The ortholog analysis was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) databse (http://www.genome.jp/kegg; release July 12, 2018).

Phylogenetic analyses

We fetched the protein sequences of TerL, head maturation protease (HMP) and the capsid from all cf-PICI in Figure S1 (Table S2). For each protein we searched for homologs among a database of all (3725) complete phages present in NCBI RefSeq (last accessed March 2021), using Diamond v2.0.6.144[37] with parameters *–query-cover 50 –ultra-sensitive –forwardonly*. We picked hits in phages that match either a capsid, HMP or a TerL of any cf-PICI with an e-value of at most $1e^{-10}$, and made a multiple alignment of these with the corresponding cf-PICI proteins, using mafft-linsi v7.490 (default parameters).[38] The multiple alignments of TerL, HMP and capsid were then purged from non-informative sites using clipkit v1.3.0[39] with the option *-m kpi-gappy*. We used the resulting multiple alignments to make phylogenetic reconstructions using maximum likelihood with iqtree v1.6.12,[40] using parameters -nt 6 -bb 1000 (to quantify the robustness of the topology) -m TEST (to find the best model). The trees were visualised with iTol v6.5.2[33] and the phage branches were collapsed for clarity.

Weighted Gene Repertoire Relatedness

We searched for sequence similarity between all proteins of all phages using mmseqs2 (release 13-45111)[41] with the sensitivity parameter set at 7.5. The results were converted to the blast format and we kept for analysis the hits respecting the following thresholds: e-value lower than 0.0001, at least 10% identity, and a coverage of at least 50% of the proteins. The hits were used to retrieve the bi-directional best hits between pairs of phages, which were used to compute a score of gene repertoire relatedness weighted by sequence identity:

$$wGRR = \frac{\sum_{i}^{p} id(Ai, Bi)}{min(A, B)}$$

where A_i and B_i is the pair *i* of homologous proteins present in *A* and *B*, id(A_i,B_i) is the sequence identity of their alignment, and min(*A*,*B*) is the number of proteins of the genome encoding the fewest proteins (*A* or *B*). wGRR is the fraction of bi-directional best hits between two elements weighted by the sequence identity of the homologs. It varies between zero (no bi-directional best hits) and one (all genes of the smallest genome have an identical homolog in the largest genome). wGRR integrates information on the frequency of homologs and sequence identity. For example, when the smallest genome has 10 proteins, a wGRR of 0.2 can result from two homologs that are strictly identical or five that have 40% identity.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification analyses were performed using Excel for Table S4. The data are presented as values with standard deviation (as mean ± SD). Statistical analyses were performed using GraphPad Prism v.9 for Figures 4B–4F, 6A, S4C, and S4D. The data are presented as values with standard deviation (as mean ± SD). For all of them, a one-way ANOVA with Dunnett's multiple comparisons test was performed to compare mean differences between wild type and individually point-mutations. Adjusted p values as: ns>0.05; *p ≤0.05; **p ≤0.01; ***p ≤0.001; ****p ≤0.0001. The reported ''n'' represents the number of experiments performed; it is detailed in the figure legends. Experiments were repeated at least three times (except for measurements of electron microscopy images that was performed two times), with sample sizes indicated in the figure legends.