nature microbiology

LETTERS

https://doi.org/10.1038/s41564-019-0612-5

A jumbo phage that forms a nucleus-like structure evades CRISPR–Cas DNA targeting but is vulnerable to type III RNA-based immunity

Lucia M. Malone[1], Suzanne L. Warring[1], Simon A. Jackson[1,2], Carolin Warnecke[1,5], Paul P. Gardner[2,3], Laura F. Gumy[4] and Peter C. Fineran[1,2]*

CRISPR–Cas systems provide bacteria with adaptive immunity against bacteriophages[1]. However, DNA modification[2,3], the production of anti-CRISPR proteins[4,5] and potentially other strategies enable phages to evade CRISPR–Cas. Here, we discovered a *Serratia* jumbo phage that evades type I CRISPR–Cas systems, but is sensitive to type III immunity. Jumbo phage infection resulted in a nucleus-like structure enclosed by a proteinaceous phage shell—a phenomenon only reported recently for distantly related *Pseudomonas* phages[6,7]. All three native CRISPR–Cas complexes in *Serratia*—type I-E, I-F and III-A—were spatially excluded from the phage nucleus and phage DNA was not targeted. However, the type III-A system still arrested jumbo phage infection by targeting phage RNA in the cytoplasm in a process requiring Cas7, Cas10 and an accessory nuclease. Type III, but not type I, systems frequently targeted nucleus-forming jumbo phages that were identified in global viral sequence datasets. The ability to recognize jumbo phage RNA and elicit immunity probably contributes to the presence of both RNA- and DNA-targeting CRISPR–Cas systems in many bacteria[1,8]. Together, our results support the model that jumbo phage nucleus-like compartments serve as a barrier to DNA-targeting, but not RNA-targeting, defences, and that this phenomenon is widespread among jumbo phages.

CRISPR–Cas systems consist of array(s) with invader-derived spacers separated by short repeats and the Cas (CRISPR-associated) proteins that provide the enzymatic machinery for immunity[1]. During phage infection, invading genetic material can be acquired into CRISPR arrays as new spacers[9,10]. Expression and processing of the CRISPR array(s) results in CRISPR RNA (crRNA) guides, onto which Cas proteins assemble, forming surveillance complexes[11,12]. In the interference step, recognition of foreign genetic material complementary to the crRNA leads to degradation of the phage nucleic acids and infection is arrested[13]. CRISPR–Cas systems are classified into two classes and six different types and bacteria often harbour multiple systems[8]. For example, *Serratia* sp. ATCC 39006 encodes three systems—type I-E and I-F, which target DNA, and type III-A, which targets DNA and RNA[14].

To identify CRISPR–Cas evasion mechanisms, we isolated phages infecting *Serratia* and assessed their sensitivity to CRISPR–Cas immunity. Of these phages, a member of the Myoviridae family was selected for further characterization and named PCH45 (Fig. 1a). PCH45 has a circularly permuted double-stranded DNA genome of 212,807 kb and is therefore a jumbo phage (that is, phages with genomes >200 kb; Fig. 1b) (ref. [15]). Sequence analysis of individual genes or the complete genome revealed that PCH45 is highly divergent from known jumbo phages, including the well-characterized *Pseudomonas* phages of the *Phikzlikevirus* genus (Fig. 1c) (refs. [16,17]). Indeed, its closest relatives, *Erwinia* phage phiEaH1 and *Serratia* phage 2050HW, showed little sequence conservation (Extended Data Fig. 1a). In summary, we identified a unique phage distinct from other described jumbo phages.

To test if the *Serratia* type I CRISPR–Cas systems elicited jumbo phage resistance, we generated strains with chromosomal anti-*gp033* (major capsid) spacers in CRISPR1 (I-E) and CRISPR2 (I-F) (Fig. 1d, Extended Data Fig. 1b and Supplementary Table 4). These anti-PCH45 spacers failed to provide jumbo phage resistance on plate or liquid cultures (Fig. 1e,f and Extended Data Fig. 1c,d), despite interfering strongly with plasmids (10^5-fold reduction in conjugation efficiency) (Fig. 1g). Importantly, these *Serratia* type I systems provided resistance against other phages, including siphovirus JS26 (ref. [18]). The jumbo phage appeared to avoid type I immunity in an anti-CRISPR-independent manner, since no known *acr* genes were detected in the PCH45 genome. Furthermore, the jumbo phage DNA was sensitive to digestion by restriction enzymes and no genes encoding known DNA modification enzymes were detected in the genome, indicating that DNA modification was not obstructing CRISPR–Cas defence (Extended Data Fig. 1i). It was possible that CRISPR–Cas expression was insufficient to provide resistance; however, the jumbo phage still evaded type I immunity when crRNAs were overexpressed from mini-CRISPR arrays (Fig. 1h–j and Extended Data Fig. 1e–h). We next examined type III immunity by expressing a spacer targeting the jumbo phage major capsid messenger RNA. In contrast to the type I systems, the type III-A system provided robust phage resistance (Fig. 1k,l). Thus, the type III-A system protected against the *Serratia* jumbo phage, whereas the type I systems were evaded via an unknown process.

We were interested in how the jumbo phage evaded type I, yet was susceptible to type III immunity. Recently, three *Pseudomonas* phages of the *Phikzlikevirus* genus were shown to produce nucleus-like structures during infection[6,7]. The phage nucleus is surrounded by a shell of phage proteins and is positioned in the cell centre by a phage-encoded tubulin spindle (PhuZ)[7,19,20]. The *Serratia* jumbo phage, despite bearing little similarity to the *Phikzlikevirus* genus, encodes a tubulin homologue (Gp187) and a potential shell protein (Gp202) (Fig. 1b). These proteins have low sequence identity

[1]Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand. [2]Genetics Otago, University of Otago, Dunedin, New Zealand. [3]Department of Biochemistry, University of Otago, Dunedin, New Zealand. [4]Department of Anatomy, University of Otago, Dunedin, New Zealand. [5]Present address: Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. *e-mail: peter.fineran@otago.ac.nz

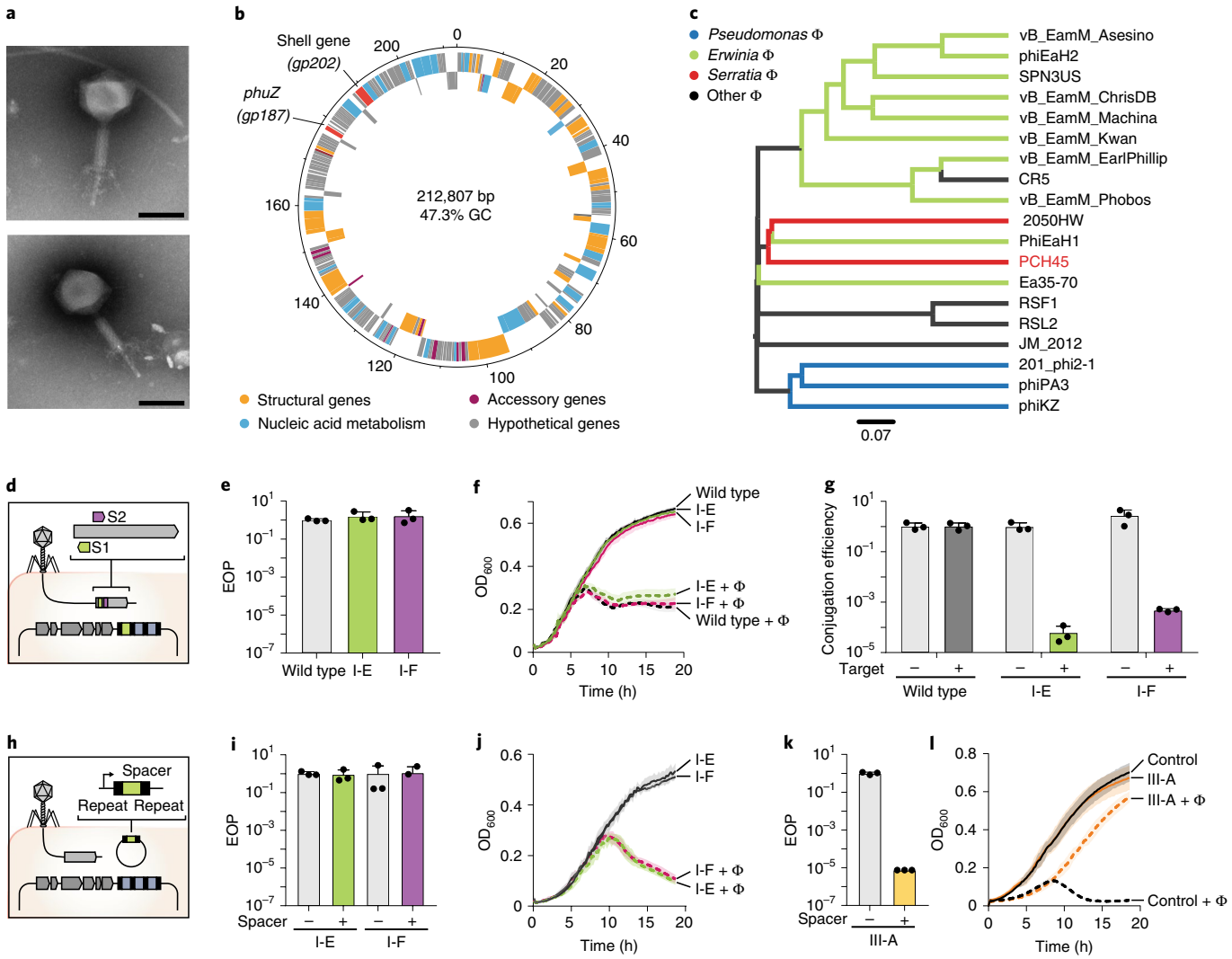


Fig. 1 | Jumbo phage PCH45 evades type I, but not type III, CRISPR–Cas immunity. a, Electron micrographs of PCH45 (scale bars, 100 nm) are representative images from three biologically independent analyses. **b**, PCH45 genome; genes on the positive (outer circle) or negative (inner circle) strand are indicated. Structural genes (orange), nucleic acid metabolism (cyan), accessory genes (purple) and hypothetical genes (grey). The *phuZ* (*gp187*) and *shell* (*gp202*) genes are shown in red. Percentage of GC refers to the guanine–cytosine content. **c**, Phylogenomic Genome BLAST Distance Phylogeny tree of jumbo phages. Branch lengths are scaled according to the distance formula (d_6). **d**, Major capsid gene (*gp033*) targets of chromosomal type I-E (S1) and I-F (S2) spacers. **e**, Phage resistance by EOP for strains with no spacers (wild type; control), type I-E spacers (S1, PCF592) or I-F spacers (S2, PCF574). **f**, Plate reader assays for strains with no spacers (wild type; control), type I-E spacers (S1, PCF592) or I-F spacers (S2, PCF574). **g**, Conjugation efficiency of untargeted (pPF1123) and targeted (*gp033*, PPF1443) plasmids into strains in **e** and **f**. **h**, Schematic of spacers expressed from mini-CRISPR arrays. **i**, Phage resistance by EOP for strains with a type I-E (S1, pPF1459) or I-F (S2, pPF1462) spacer in mini-CRISPR arrays. **j**, Plate reader assays for strains with a type I-E (S1, pPF1459) or I-F (S2, pPF1462) spacer in mini-CRISPR arrays. For **i** and **j**, plasmids lacking spacers (I-E: pPF974 and I-F: pPF975) were the controls . **k**, Phage resistance by EOP for strains with a type III-A (S3, pPF1467) spacer in a mini-CRISPR array. **l**, Plate reader assays for strains with a type III-A (S3, pPF1467) spacer in a mini-CRISPR array. For **k** and **l**, a mini-CRISPR array lacking spacers (pPF975) was the control. **e**–**g**,**i**–**l**, Data are presented as the mean ± s.d. ($n = 3$ biologically independent samples). **f**,**j**,**l**, Multiplicity of infection (MOI) = 0.001.

(16.5 and 19.9% at the amino acid level for PhuZ and the shell protein, respectively) to those in the *Pseudomonas* phage phiKZ (type species of the *Phikzlikevirus* genus) (Extended Data Fig. 2a,b). Therefore, we hypothesized that the *Serratia* jumbo phage produces a nucleus-like compartment on infection. *Serratia* was infected with PCH45 and confocal microscopy was used to visualize DNA and membranes. During infection, we observed circular DNA foci, consistent with nucleus-like structures (Fig. 2a). By contrast, DNA was evenly distributed in uninfected controls. Thirty minutes after infection, most phage nuclei were either localized centrally ($n = 102$; 61%), or towards the cell poles ($n = 64$; 39%). To test if the DNA foci were encapsulated by a phage shell, the putative shell protein was tagged (monomeric enhanced green fluorescent protein (mEGFP)-Gp202). On phage infection, the tagged shell protein assembled into a spherical structure enclosing the phage DNA but no shell was formed without infection (Fig. 2b). Therefore, phage infection leads to DNA accumulation within a phage-encoded protein shell that includes protein Gp202.

In *Pseudomonas Phikzlikevirus* 201Φ2-1, shell formation allows the selective translocation of proteins into the phage nucleus, restricting other proteins to the cytoplasm[6]. We hypothesized that the *Serratia* jumbo phage evades DNA targeting due to the nucleus-like compartment excluding Cas proteins from phage DNA. Therefore, we monitored CRISPR–Cas interference complex localization during jumbo phage infection with the large subunit of all systems tagged with mCherry2 (I-E, *cas8e*; I-F, *cas8f*; III-A,

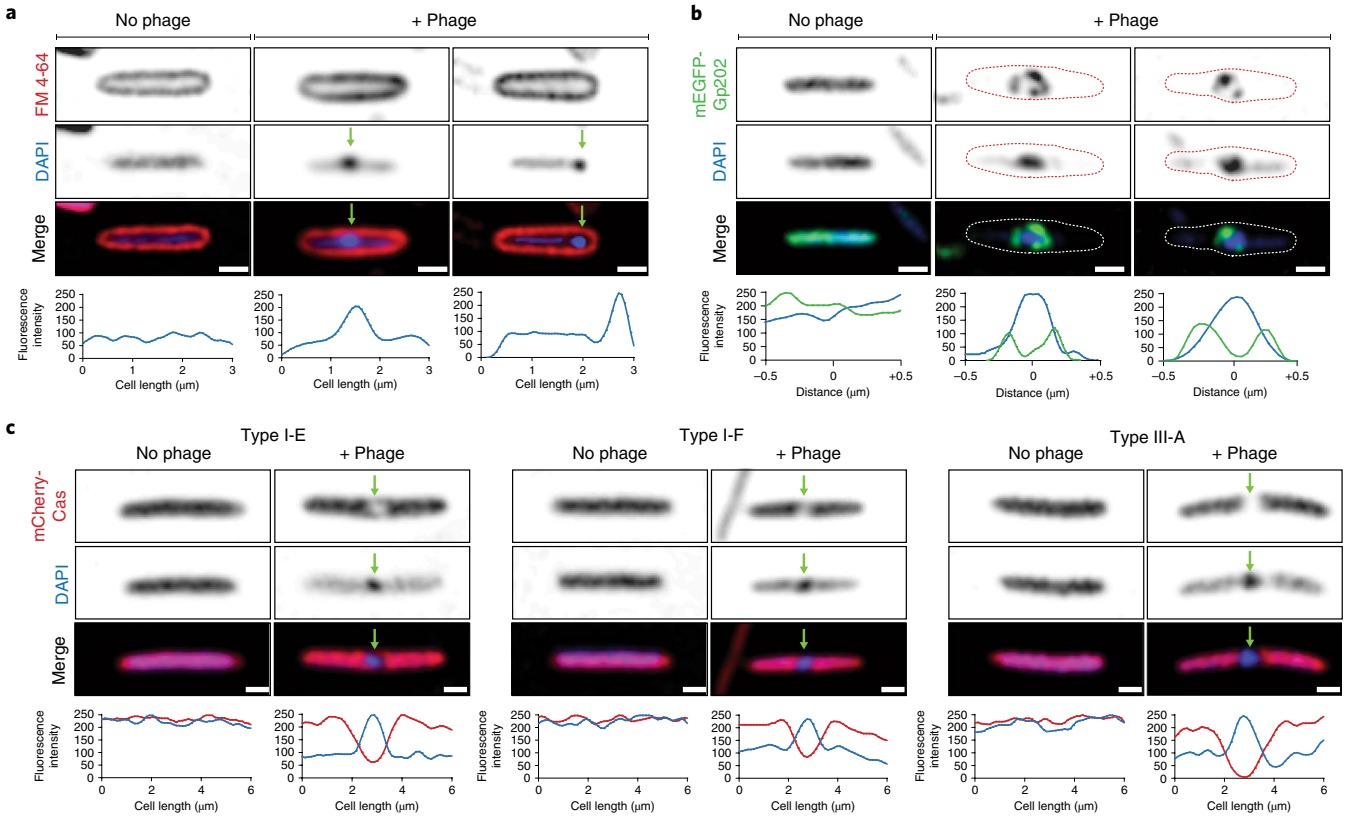


Fig. 2 | The jumbo phage assembles a DNA-containing protein shell during infection that excludes CRISPR–Cas complexes. a, Nucleus-like structures (arrows) form in *Serratia* wild-type infected with PCH45 (+ phage). Uninfected cells (no phage) do not form DNA foci. DNA (blue) and membranes (red) were stained with DAPI and FM 4-64, respectively. Quantifications show the fluorescence intensity distribution of DAPI along the cell length of the representative single cells. Scale bars, 1 μm. **b**, Gp202 forms a shell (green) around the DNA foci (blue). Quantifications show the fluorescence intensity profile of the DNA and shell in the representative single cells. The dashed lines outline the cell shape. Scale bars, 1 μm. **c**, Cas type I-E, I-F and III-A complexes (red) are excluded from DNA (blue) in infected cells (+ phage) (arrows). Scale bars, 1 μm. The graphs show the fluorescence intensity distributions of the Cas complexes (red) and DNA (blue) across the cell length of single cells. Images were collected 30 min post-infection (MOI = 8) and are representative of the typical phenotype observed for infected cells from three biologically independent experiments.

cas10). All tagged systems retained interference activity against the CRISPR-sensitive phage JS26 (Extended Data Fig. 2c). For all CRISPR–Cas types, the interference complexes were localized in the cytoplasm, external to the phage nucleus (Fig. 2c). These results, combined with the absence of known *acr* and DNA modification genes, together with the phage genomic DNA (gDNA) degradation by several restriction enzymes, suggest that the shell protects the phage DNA by spatial exclusion of the CRISPR–Cas complexes.

Replication and transcription of *Phikzlikevirus* DNA occurs inside the nucleus-like compartment and mRNA is transported to the cytoplasm for translation[6]. This is consistent with the *Serratia* jumbo phage being insensitive to type I systems (target DNA), while remaining sensitive to type III (targets both RNA and DNA)[21,22]. To further investigate the role of type III RNA targeting in jumbo phage defence, we tested a panel of crRNAs that target different PCH45 genes (Fig. 3a). In agreement with RNA targeting, all antisense crRNAs inhibited jumbo phage infection in plate and liquid assays, whereas sense crRNAs to transcribed mRNAs provided no protection (Fig. 3b–d). Targeting was unaffected by whether the target RNA was predicted to be expressed in an early, middle or late stage of infection (Fig. 3a,b).

In type III-A systems, RNAs are recognized by the Csm effector complex[21,22], triggering sequence-specific RNA cleavage by Cas7 (ref. [23]). In addition, the nuclease (HD) domain in Cas10 promotes non-specific DNA cleavage[24]. RNA binding also activates the Cas10 Palm domain, which synthesizes cyclic oligoadenylate (cOA) secondary messengers that activate non-specific RNases (Csm6 in the type III-A systems studied), causing collateral RNA degradation[25–27]. This collateral cleavage of RNA has been proposed to induce cell dormancy that may prevent phage proliferation[28]. However, the precise mechanisms of type III-A-mediated phage resistance remains poorly understood. The *Serratia* type III-A system has two CRISPR arrays (CRISPR3 and CRISPR4), an operon encoding the adaptation complex and an operon encoding the Csm complex (Fig. 3e). In addition, a hypothetical nuclease is convergently transcribed between *cas6* and CRISPR4. This nuclease is unrelated to Csm6, but has similarity to NucC, which an unpublished report indicates is activated by cOA signals and non-specifically degrades DNA in vitro[29]. To investigate jumbo phage protection by the *Serratia* type III-A system, catalytic mutants of the key proteins involved in RNA and DNA cleavage were tested with multiple spacers (Fig. 3f,g and Extended Data Fig. 3a). As predicted, a Cas10 HD mutation (*cas10*$^{H17A,\ N18A}$) did not affect type III-A immunity, indicating that DNA cleavage by Cas10 is not necessary for jumbo phage resistance. In contrast, active site mutations in Cas7 (also known as Csm3; *csm3*^{D34A}) (ref. [23]) or the Cas10 Palm domain (*cas10*$^{D618A,\ D619A}$), which disrupts cyclic oligoadenylate signalling[26], abolished phage resistance. Moreover, resistance was also lost when the accessory nuclease was deleted. The same effects on interference of plasmid conjugation were observed for all mutants and restoration of the wild-type copy of each mutant gene complemented CRISPR–Cas activity (Extended Data Fig. 3b,c). To further

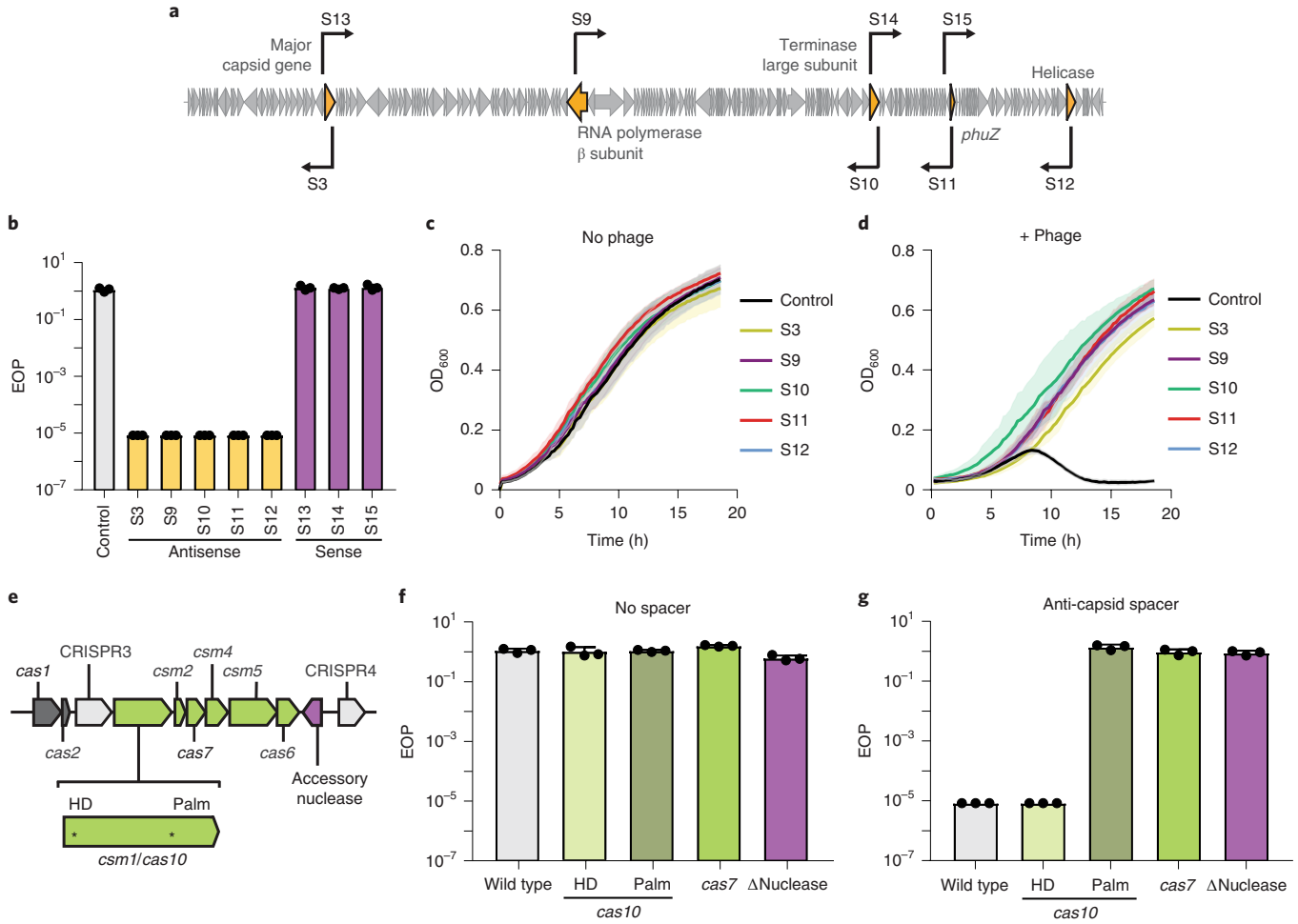


Fig. 3 | Jumbo phage infection is inhibited by the type III CRISPR–Cas system. a, PCH45 phage genome indicating gene targets of the type III-A spacers: major capsid (*gp033*; S3 and S13), RNA polymerase β subunit (*gp084*; S9), terminase large subunit (*gp159*; S10 and S14), tubulin-like protein (*gp187*; S11 and S15) and helicase (*gp217*; S12). **b**, EOP assay for strains expressing type III-A spacers from mini-CRISPR arrays (control; pPF975, S3; pPF1467, S9; pPF1466, S10; pPF1469, S11; pPF1470, S12; pPF1468, S13; pPF1994, S14; pPF1995, S15; pPF1996). **c**, Plate reader assay of uninfected strains expressing the spacers. **d**, Plate reader assay of strains infected with PCH45 (MOI = 0.001). **e**, *Serratia* type III-A locus. Adaptation genes (grey), CRISPR arrays (light grey), interference genes (green) and accessory nuclease (pink) are shown. **f**, EOP assay with no spacer (pPF975). **g**, EOP assay with an anti-capsid spacer (S3) expressed from a mini-CRISPR array in the wild type or type III-A mutants: *cas10*$^{H17A, N18A}$ (HD), *cas10*$^{D618A, D619A}$ (Palm); *cas7*^{D34A}; and the nuclease deletion. **b**–**d**,**f**,**g**, Data are presented as the mean ± s.d. ($n = 3$ biologically independent samples).

understand phage protection provided by the *Serratia* type III-A system, it is necessary to determine the mechanism of the accessory nuclease, which we predict responds to cOA signals, and whether induction of dormancy is involved in immunity. We have shown that jumbo phage immunity requires the specific RNA-targeting and cyclic oligonucleotide-signalling capabilities of the type III-A CRISPR–Cas system and requires the accessory nuclease.

We hypothesized that type III CRISPR–Cas immunity against nucleus-forming jumbo phages would occur in natural environments, whereas type I immunity would be rare. To test this hypothesis, we analysed type I-E, I-F and III spacers from approximately 160,000 bacterial genomes and identified their targets in isolated phage genomes and viral contigs from global metagenomes (Supplementary Table 2 and Extended Data Fig. 4a–d). For total phages, many targets were identified for all systems (Fig. 4a). Consistent with our model, targets of type III spacers were significantly enriched in jumbo phages (that is, >200 kb; $\chi^2 < 0.001$) (Fig. 4a,b) and further enriched in nucleus-forming phages that encode homologues of both shell and tubulin proteins ($\chi^2 < 0.001$) (Fig. 4c). Multiple examples of type III systems targeting nucleus-forming jumbo phages were present in diverse classes of proteobacteria. In contrast, type I-E and I-F spacer matches were depleted in jumbo phages and those defined as nucleus-forming (Fig. 4b,c and Extended Data Fig. 4e). The presence of spacers in type III arrays that target jumbo phages raises the question of how they are acquired, which requires future work. RNA may be the spacer source via reverse transcriptase-Cas1 fusion proteins[30] or assisted by other reverse transcriptases. DNA could be a substrate for adaptation before protection by the nucleus or when injected from defective phages[31]. Finally, adaptation complexes might be able to access DNA in the nucleus. In conclusion, both experimental and bioinformatics data provide evidence that type III CRISPR–Cas immunity against jumbo phages is widespread in nature, and that these phages evade type I immunity.

We have discovered a jumbo phage that evades DNA targeting by two native type I CRISPR–Cas systems while retaining sensitivity to the RNA-targeting capabilities of the type III-A system. We propose that resistance is conferred by the formation of a nucleus-like structure in the bacterial cytoplasm that physically shields DNA, but not RNA, from cytoplasmic CRISPR–Cas effector complexes.

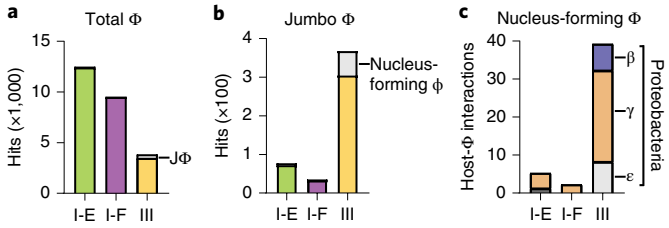


Fig. 4 | Jumbo phages are targeted by type III CRISPR–Cas systems in different bacterial classes. a, Number of spacers in type I-E, I-F or type III systems matching total *Caudovirales* phages, with jumbo phages (JΦ) in grey. **b**, Number of spacers in type I-E, I-F or type III systems matching jumbo phages with nucleus-forming phages in grey. **c**, Number of spacers in type I-E, I-F or type III systems matching unique host nucleus-forming phage interactions.

This concept of exclusion defence was proposed earlier for phages of the *Phikzlikevirus* genus[7] and is supported by a recent unpublished study showing that phage DNA, but not RNA, is protected from immune systems different to those tested in this study[32]. Our work on a unique jumbo phage that infects a different order of bacteria, and bears little similarity to the *Phikzlikevirus* genus, coupled with our bioinformatic analyses, provides evidence that the phage nucleus is a widespread counter-defence strategy among jumbo phages. This manner of immune evasion leads to the prediction that the phage nucleus would provide broad protection from diverse DNA-targeting defence systems. This quality makes nucleus-forming phages prime candidates for phage-based therapies. Importantly, despite DNA protection, RNA export to the cytoplasm is a vulnerability of jumbo phages that can be exploited by type III CRISPR–Cas systems. Jumbo phage infection has probably played a role in selecting for the observed widespread type III RNA-targeting immunity in strains already possessing DNA-based defences.

Methods

Bacterial strains, plasmids and culture conditions. The plasmids and primers used in this study are listed in Supplementary Tables 3 and 4 and bacterial strains are listed in Supplementary Table 5. Plasmids were all confirmed by sequencing, transformed into *Escherichia coli* ST18 and introduced into *Serratia* strains by conjugation. *Serratia* sp. ATCC 39006 (*Serratia*) and *E. coli* were grown at 30 and 37 °C in lysogeny broth under shaking conditions (160 r.p.m.). When grown on plates, containing 1.5% w/v agar in lysogeny broth, the strains were incubated at an appropriate temperature until colonies appeared. Media were supplemented with ampicillin (100 μg ml^{-1}), chloramphenicol (25 μg ml^{-1}), kanamycin (50 μg ml^{-1}), tetracycline (1 μg ml^{-1}) and 5-aminolevulinic acid (50 μg ml^{-1}) when required.

Jumbo phage isolation. Jumbo phage PCH45 was isolated from sewage samples extracted from the Tahuna Waste Water Treatment Plant in Dunedin, New Zealand (45° 54′ 16.1 S; 170° 31′ 16.8 E). Enrichment for phages infecting *Serratia* was performed by mixing 100 μl of sewage sample with 5 ml of *Serratia* culture and incubating the mixture overnight at 30 °C, shaking at 160 r.p.m. The phage suspension was separated from the cellular debris by centrifugation (2,000*g* for 20 min at 4 °C), diluted and plated onto *Serratia* lawns. Jumbo phages were selected based on their small plaque morphology, picked and used to infect a *Serratia* culture overnight. Serial dilutions of the lysate were plated onto lysogeny broth agar (LBA) plates and grown overnight. The process was repeated until uniform plaques were obtained to ensure a pure phage stock.

Preparation of phage stocks and titration. Phage stocks were prepared as described elsewhere[33]. Briefly, 100 μl of overnight bacterial culture was mixed with serial dilutions of phage lysate and added to 4 ml of LBA overlay (0.35% w/v), which was then poured onto LBA plates. Plates were then incubated overnight at 30 °C; plates with webbed lysis (almost confluent lysis) had the top agar scraped off and pooled together into a centrifuge tube. A few drops of chloroform ($NaCO_3$-saturated) were added before thoroughly vortexing the mixture to lyse the cells. Finally, a centrifugation step was performed (2,000*g* for 20 min at 4 °C) to separate the virions from the cell debris. The supernatant was placed in a sterile universal for storage and phage titre was determined by pipetting 20 μl drops of serial dilutions of the phage stock onto a LBA overlay seeded with 100 μl of *Serratia* overnight culture. Plaques were counted after incubation overnight at 30 °C, with the phage titre represented as plaque-forming units (PFUs) per millilitre (PFU ml^{-1}). Phage stocks were stored at 4 °C.

Electron microscopy. To examine the jumbo phage by transmission electron microscopy, 10 μl of high-titre phage stock (approximately 10^9 PFU ml^{-1}) was loaded onto plasma-glowed, carbon-coated 300 mesh copper grids. After 60 s, the excess specimen was removed by blotting and 10 μl of 1% w/v phosphotungstic acid (pH 7.2) was applied to the grid to stain the samples and blotted off immediately after. The grids were viewed in a CM100 BioTWIN transmission electron microscope (Philips/FEI Corporation) and images captured using a MegaView lll digital camera (Soft Imaging System).

Phage DNA extraction and restriction analysis. DNA was extracted from a high-titre phage stock (approximately 10^{10} PFU ml^{-1}), using the cetyltrimethylammonium bromide (CTAB) method described elsewhere[34]. Briefly, 2.5 ml of phage stock were treated with 100 ng of RNase A and 100 U of DNAse I and incubated at 37 °C for 30 min to clean the phage from bacterial nucleic acid. Enzyme inactivation was performed by adding 0.8 ml of 0.58 M EDTA (pH 8.0). Next, phage proteins were removed by adding proteinase K (20 mg ml^{-1}) and incubating at 45 °C for 15 min. CTAB (10% w/v) in 4% NaCl (w/v) was preheated to 55 °C, added to the sample, mixed gently by inversion and cooled on ice for 15 min. The CTAB–DNA complex was collected by centrifugation (12,000*g* for 15 min) and the pellet was resuspended in 1 ml of 1.2 M NaCl. Phage DNA was precipitated by adding 600 μl of isopropanol, washed with 75% ethanol and resuspended in 100 μl of Tris-EDTA buffer. Samples were cleaned with the DNeasy Blood & Tissue Kit (QIAGEN) according to the manufacturer's instructions. DNA concentration was determined by fluorometric quantification using the Qubit dsDNA HS Assay Kit and the Qubit Fluorometer v.2.0 (Invitrogen) according to the manufacturer's instructions. Restriction digestion was performed on PCH45 DNA to check for DNA modification. Digestion of 1 μg of PCH45 genomic DNA was performed with different restriction enzymes (New England Biolabs). The samples were incubated for 2 h at 37 °C and separated on a 1% agarose gel alongside the 1 kb Plus DNA Ladder (Thermo Fisher Scientific), stained with ethidium bromide and visualized under ultraviolet light.

Genome sequencing, annotation and comparative genomics. Library preparation and sequencing was performed by the Massey Genome Service (Massey University). First, DNA libraries were prepared with the Nextera XT DNA Library Preparation Kit v.2 (Illumina). Quality control was checked using the Quant-iT dsDNA HS Assay for quantification and analysed using SolexaQA++, FastQC and FastQ screen. Sequencing was performed on an Illumina MiSeq System (2 × 150 bp); the resulting reads were processed and trimmed using the SolexaQA++ software v.3.1.7.1. The resulting reads were assembled de novo using SPAdes v.3.9.0 (ref. [35]) and annotated with the RAST tool kit[36]. Further annotation was performed manually by analysing the protein Basic Local Alignment Search Tool (BLASTp v.2.10.0) hits for each independent coding sequence identified by RAST tool kit v.2.0. tRNAscan-SE v.2.0 (ref. [37]) was used to identify putative transfer RNAs. The genome was visualized using DNAplotter v.17.0.1 (ref. [38]). Similar phage genomes were identified using pairwise sequence comparison[39] and visualized using Easyfig v.2.2.2 (ref. [40]). Finally, the taxonomic trees were built using Virus Classification and Tree building Online Resource[41] and modified with FigTree v.1.4.3.

Generation of native type I-E and I-F anti-phage strains. *Serratia* strains harbouring anti-PCH45 spacers in their type I-E and I-F chromosomal arrays were obtained by primed spacer acquisition[42]. Plasmids primed by the type I-E and I-F systems (pPF1255 and pPF1256) were generated as follows. The major capsid gene (*gp033*) was amplified by PCR from phage gDNA using PF2231/PF2232. The insert was digested with SpeI and KpnI (New England Biolabs) and cloned into the priming vectors (pPF1125 and pPF1126) that were previously digested with the same enzymes. pPF1125 and pPF1126 contained a protospacer primed by spacer 1 from either the type I-E or I-F CRISPR–Cas systems in *Serratia*. Plasmids were transformed into *E. coli* ST18, plated onto LBA + 5-aminolevulinic acid + chloramphenicol and grown overnight at 37 °C. The vectors were conjugated into *Serratia* by mixing equal volumes of donor and recipient and plating the mating spot onto LBA + 5-aminolevulinic acid. The mating spot was streaked onto LBA + chloramphenicol to select for *Serratia* transconjugants. *Serratia* colonies were grown overnight in the absence of antibiotic to allow plasmid loss. Cultures were passaged for 3 d by inoculating lysogeny broth with 5 μl of overnight culture. Each day, serial dilutions were plated and incubated at 30 °C until colonies appeared on the plate. The clones were screened by PCR in search of CRISPR array expansion (PF1989–PF1887 for CRISPR1 and PF1990–PF1889 for CRISPR2). Where CRISPR expansion was observed, PCR products were sequenced to determine if the spacers acquired corresponded to the plasmid backbone or the phage fragment (Supplementary Table 3 and Extended Data Fig. 5). The clones were also patched onto LBA + chloramphenicol to check for plasmid loss.

Plasmid expression of anti-phage spacers. For some experiments, spacers were expressed from plasmid mini-CRISPR arrays. These plasmids were constructed by cloning anti-PCH45 and anti-JS26 spacers into plasmid mini-CRISPR arrays.

Briefly, reverse complement primers carrying spacer sequences targeting phage genes flanked by two BsaI restriction sites were annealed (details in Supplementary Tables 3 and 4 and Extended Data Fig. 6). Annealed primers were cloned into plasmids pPF974, pPF975 (ref. [18]) and pPF976 (expression plasmids carrying type I-E, I-F and III-A mini-CRISPR arrays) previously digested with BsaI. pPF976 was constructed in the same manner as pPF974/5 (ref. [18]) using the PF1981 and PF1982 primers. The plasmids carrying the mini-CRISPR arrays were introduced into *E. coli* ST18 by transformation and then conjugated into *Serratia* (and derivative mutants). Expression of crRNAs was induced by adding isopropyl-β-D-thiogalactoside (IPTG) (0.1 mM) and strains were grown in the presence of kanamycin for plasmid maintenance.

Phage resistance efficiency of plaquing assay. To assess the infectivity of PCH45 and JS26 on different *Serratia* strains, efficiency of plaquing (EOP) assays were performed. A soft LBA overlay (0.35% w/v) containing 100 µl of bacterial culture was poured onto a LBA plate. Serial tenfold dilutions of high-titre phage stock (approximately 10^9 PFU ml^{-1}) were spotted (20 µl) onto the agar overlay and plates were incubated overnight at 30 °C. The EOP was calculated as the ratio of PFU ml^{-1} produced on tested strains to the PFU ml^{-1} on the control *Serratia* strain. All conditions were repeated in triplicate and plotted as the mean ± s.d.

Phage resistance infection time courses. *Serratia* cultures were grown from an initial optical density OD_{600} = 0.05 at 30 °C, shaking (160 r.p.m.) until reaching an OD_{600} = 0.3 (exponential phase). Cultures were diluted to an OD_{600} = 0.05, 180 µl aliquots were pipetted into a 96-well plate and 20 µl of diluted phage lysate was added to produce a multiplicity of infection (MOI) = 0.001. The 96-well plate was incubated in a Varioskan Flash plate reader (Thermo Fisher Scientific) at 30 °C with 240 r.p.m. shaking and OD_{600} measurements taken at 12 min increments for 20 h. Each condition was repeated in triplicate with data plotted as the mean ± s.d. For strains with anti-phage spacers expressed from a plasmid, media were supplemented with kanamycin and IPTG (0.1 mM).

Construction of shell and tubulin-like protein phylogenetic trees. To build a protein phylogenetic tree, protein homologues were identified with a BLASTp search using phiKZ and PCH45 shell and PhuZ proteins as the query (PCH45_Gp202/Gp187 and phiKZ_Gp054/Gp039, respectively). Protein hits (expect value < 10^{-10}) encoded by jumbo phages were used in a CustalW alignment and a phylogenetic tree was constructed using RAxML v.8.1.20 (ref. [43]) run with the PROTGAMMAJTT model, default parameters and branch support computed out of 100 bootstrapped trees. The raw files used to generate the phylogenetic trees can be found as Supplementary Files 1 and 2.

Construction of mEGFP/shell protein expression plasmid. A plasmid (pPF1956) for expression of an mEGFP-tagged shell protein was generated by PCR amplification of the shell gene (*gp202*) using PF3825/PF3812 and mEGFP, with PF3826/PF3827 from the phage gDNA and gBLOCK PF3809, respectively. The pQE-80L-oriT vector stuffer was digested with SphI and KpnI (New England Biolabs) and assembled with the two inserts using Gibson Assembly using HiFi DNA Assembly Master Mix (New England Biolabs) following manufacturer's instructions.

Construction of fluorescently tagged Cas complexes. Cas complexes were fluorescently tagged by fusing mCherry2 to the N-terminal region of each large subunit (*cas8e*, *cas8f* and *cas10*). Plasmids were constructed by cloning mCherry2 flanked by the upstream region and the first 500 bp of each gene into the suicide vector pPF1117 (ref. [11]). The regions were amplified by PCR from gBLOCK PF3810 (mCherry2 + linker) and *Serratia* wild-type using the primers PF3817/PF3818, PF3811/PF3812, PF3819/PF3820 (mCherry2-*cas8e*, pPF1951), PF3821/PF3822, PF3811/PF3812, PF3823/PF3824, (mCherry2-*cas8f*, pPF1953) and PF3813/PF3814, PF3811/PF3812, PF3815/PF3816 (mCherry2-*cas10*, pPF1955). The inserts were cloned into pPF1117, previously digested with SphI and SalI, using Gibson Assembly. The fluorescently tagged Cas complexes were introduced into the chromosome by homologous recombination. The suicide vectors for *cas8e* (pPF1951), *cas8f* (pPF1951) and *cas10* (pPF1955) were conjugated into *Serratia*, recombination was selected by growth on LBA with 20% w/v sucrose and colonies were screened for chloramphenicol sensitivity as described elsewhere[44,45]. This gave strains PCF734 (mCherry-*cas8e*), PCF736 (mCherry-*cas8f*) and PCF732 (mCherry-*cas10*). To optimize the expression of the fluorescent Cas complexes, an IPTG-inducible T5 promoter was inserted upstream of each complex. mCherry2 was amplified by PCR from gBLOCK PF3810 using the PF4005/PF4007 primers and cloned into pPF1813 (ref. [46]) previously digested with EcoRI and XmaI (New England Biolabs), using Gibson Assembly to yield plasmid pPF2036. Plasmid pPF2036 was inserted at the mCherry2 site by homologous recombination in PCF732, PCF734, PCF736 and the function of the complexes was confirmed in EOP assays using the target plasmids pPF1473, pPF1485 and pPF1489 against *Serratia* siphovirus JS26 as described earlier.

Confocal microscopy. To observe infected and uninfected bacteria utilizing microscopy, cultures of bacteria were grown overnight from single colonies. These were then subcultured and grown to an OD_{600} between 0.2 and 0.3 (incubated at 30 °C with 160 r.p.m. shaking). For infected samples, phage PCH45 was added to an MOI of approximately 8 with an infection time of 30 min. Aliquots of 3 ml of each culture were then pelleted and washed with minimal medium (40 mM K_2HPO_4, 14.7 mM KH_2PO_4, 0.1% $(NH_4)_2SO_4$, 0.4 mM $MgSO_4$ and 0.2% sucrose, pH 7.1) before being stained with 4,6-diamidino-2-phenylindole (DAPI; 4 µg ml^{-1}) and FM 4-64 (12 µg ml^{-1}) for 20–30 min. Stained samples were then washed twice with minimal medium before being resuspended in 100 µl of minimal medium. To suspend samples, 15 µl aliquots of each sample were mixed with 15 µl of molten 1.2% agarose (made up in minimal media) before being sealed onto microscope slides with a coverslip.

To observe the formation of a shell structure, a tagged putative shell gene (mEGFP-*gp202*) was expressed from plasmid pPF1956. Due to leaky expression of the mEGFP-*gp202* shell gene, no induction was required. To observe the localization of interference complexes on phage infection, type I-E, type I-F and type III-A complexes were tagged with mCherry2 as described earlier. Expression of mCherry2-tagged type I-E, I-F and III-A Cas complexes was induced with the addition of IPTG (10 µM) at the time of infection.

Images were acquired using a CFI Plan APO Lambda ×100 1.49 numerical aperture oil objective (Nikon Corporation) on the multimodal imaging platform Dragonfly v.505 (Oxford Instruments) equipped with 405, 488, 561 and 637 nm lasers built on a Nikon Ti2-E microscope body with Perfect Focus System (Nikon Corporation). Data were collected in Spinning Disk 40 µm pinhole mode on the iXon888 EMCCD camera with ×2 optical magnification using the Fusion Studio v.1.4 software. Z stacks were collected with 0.1 µm increments on the *z* axis using an Applied Scientific Instrumentation stage with 500 µm piezo z drive. Data were visualized using the Fiji software (Windows 64-bit). Microscopic images were further processed by the deconvolution algorithm in the Huygens Suite scientific volume imaging Image Analysis Program (Huygens Professional).

Image analysis. To quantify the fluorescence intensity distribution in single cells in Fig. 2, the length of cells and the length of nucleus-like structures were traced using the Fiji software. Subsequently, the intensity profile along the linescan was measured and plotted in Prism v.8.3.0 (GraphPad Software).

Construction of chromosomal type III-A mutants. To mutate *cas7* and the accessory nuclease in the type III-A system, these genes were first replaced in the chromosome with a kanamycin cassette by homologous recombination. To generate the knockout constructs (pPF1929; *cas7*, pPF1932; nuclease), the upstream and downstream regions of each gene were amplified by PCR using the primer pairs PF3750/PF3585, PF3754/PF3751 (*cas7*) and PF3743/PF3745, PF3748/PF3749 (nuclease) using *Serratia* DNA as a template. The kanamycin cassette was cloned from pSEVA211 (ref. [47]) using the primer pairs PF3752/PF3753 (for pPF1929) and PF3746/PF3747 (for pPF1932). The suicide vector pPF1117 (ref. [18]) was digested with SphI and SalI and the inserts were cloned using the Gibson Assembly (HiFi DNA Assembly Master Mix; New England Biolabs). To generate the *cas10* knockout vector (pPF927), the primer pairs PF1934 (SalI site)/PF1935 (BamHI site) and PF1936 (BamHI site)/PF1937 (SphI) were used to amplify by PCR the *cas10* upstream and downstream regions from *Serratia* wild-type colonies as the DNA template. The two inserts were cloned with a three-part ligation including the suicide vector pPF923 (ref. [18]) previously digested with SalI and SphI. The kanamycin^R-marked deletion strains were generated (PCF682; *cas7* and PCF685; nuclease) using the plasmids pPF1929 and pPF1932 via homologous recombination as described for the tagged Cas complex strains. A markerless *cas10* deletion (PCF303) was constructed by homologous recombination using plasmid pPF927.

Plasmids for site-directed mutagenesis of *cas10* were constructed as described further on. Point mutations were introduced by overlap extension PCR from the *Serratia* DNA template using primers carrying the altered sequence and a complementary region with an overlapping primer. Amplification was performed using the following primer pairs: PF3756/PF2167 and PF2166/PF3757 for pPF1936 (Cas10 HD domain mutant) and PF3756/PF2127 and PF2126/PF3757 for pPF1938 (Cas10 Palm domain mutant). For pPF1930 (*cas7*^{D34A}), the upstream and downstream regions of *cas7* were cloned using the primer pairs PF3750/PF3585 and PF3755/PF3751, respectively and *Serratia* wild-type colonies as the DNA template, and primer pairs PF3589/PF3590 and gBLOCK PF3591 as the DNA template. To delete the accessory nuclease, pPF1933 was constructed using the primer pairs PF3743 (overlap with pPF1117)/PF3745 (overlap with PF3749), PF3749 (overlap with PF3745)/PF3744 (overlap with pPF1117) and *Serratia* as the template DNA. All inserts were cloned into pPF1117, previously digested with SalI and SphI restriction enzymes (New England Biolabs), using Gibson Assembly.

The type III-A mutant strains PCF683 (*cas7*^{D34A}), PCF690 (*cas10*$^{H17A, N18A}$, HD mutant), PCF691 (*cas10*$^{D618A, D619A}$, Palm mutant) and PCF686 (unmarked deletion of the type III-A accessory nuclease; CWC46_RS19930) were generated by homologous recombination as described for the tagged Cas complex strains. The genes carrying the appropriate point mutations were introduced by homologous recombination with plasmids (pPF1930, pPF1933, pPF1934 and pPF1935) into the *cas7* (PCF682) and *cas10* (PCF303) deletion mutants. To generate the unmarked nuclease mutant (PCF685), pPF1933 was recombined into the marked nuclease mutant (PCF686).

Complementation of type III-A chromosomal mutants. The type III-A mutants were complemented by reinserting a wild-type copy of each of the genes into the mutant backgrounds. The constructs pPF1931 (*cas7*), pPF1934 (accessory nuclease) and pPF1935 (*cas10*) were generated using the primers PF3750/PF3751, PF3743/PF3744 and PF3756/PF3757, and cloned into pPF1117 as described earlier. Unmarked genes were reinserted by homologous recombination into strains PCF682 (Δ*cas7*), PCF685 (Δnuclease) and PCF303 (Δ*cas10*) with plasmids carrying wild-type genes (pPF1931; *cas7*, pPF1934; nuclease and pPF1935; *cas10*).

Type I-E and I-F plasmid interference assay. To generate a vector targeted by the anti-PCH45 type I-E and I-F spacers (pPF1443), the priming protospacer in pPF1255 was removed. After digestion with SpeI and SphI, pPF1255 was gel-extracted, treated with Mung Bean Nuclease and re-ligated using T4 DNA ligase. The interference ability of *Serratia* strains (PCF592 and PCF547) carrying type I-E and I-F anti-PCH45 spacers was tested in conjugation efficiency assays. Cultures of donor *E. coli* ST18 carrying plasmids pPF1123 (untargeted) or pPF1443 (targeted) and the *Serratia* recipient strains were grown overnight. Cultures were adjusted to an $OD_{600} = 1$ and donor and recipient strains were mixed in a 1:1 ratio. The mixture was spotted onto LBA + 5-aminolevulinic acid and incubated overnight at 30 °C. Next, the mating spots were scraped from the plate and resuspended in 1 ml of PBS. Tenfold serial dilutions were performed and 10 µl of each were spotted onto LBA (total recipient count) and LBA + chloramphenicol (transconjugants). Conjugation efficiency was calculated as transconjugants (colony-forming units per millilitre (CFU ml^{-1}))/total recipients (CFU ml^{-1}). For the type III-A mutants (PCF683, PCF686, PCF690, PCF691) and their complemented controls (PCF684, PCF687, PCF688), plasmids pPF781 (untargeted) and pPF1043 (targeted) were used. Conjugation efficiency was performed as described for the type I interference assay, but with the following differences. LBA plates for the mating spots included glucose (0.2% w/v) and arabinose (0.02% w/v) was included in all plates for transconjugant and total recipient counts to induce protospacer (target site) transcription.

Identification of phages targeted by type III spacers. Type III, I-E and I-F CRISPR–Cas hosts were identified by the presence of *cas* gene annotations in the RefSeq release 95 Bacterial genome database. CRISPR loci were extracted using CRISPRDetect v.2.4 (ref. [48]) with a cut-off score of 2.5. For predicted type III systems, we excluded CRISPRs with repeats that matched known non-type III systems[49] or with repeat lengths <30 nucleotides (nt) or >50 nt, and all spacers <25 nt or >45 nt. For type I-E systems, repeats <28 nt or >32 nt, and spacers <28 nt or >34 nt were excluded. For type I-F systems, repeats <26 nt or >30 nt, and spacers <28 nt or >34 nt were excluded. We searched for matches to the spacers against *Caudovirales* genomes in GenBank and non-eukaryotic viral contigs in the IMG/VR database v.4 (July 2018) (ref. [50]) using the Global Alignment Short Sequence Search Tool v.1.2.8 (ref. [51]) (seed = 8, sensitivity = 3, match = 80, gaps = 0). Spacer-target matches were scored along their full-length alignment as +1 match and −1 mismatch. The dinucleotide-shuffled control datasets, used to determine appropriate scoring cut-offs for the spacer-target matches (Extended Data Fig. 4a–d), were generated using fasta-shuffle-letters from the MEME Suite v.5.05 (ref. [52]). Redundant spacer-target matches, due to similar host CRISPRs or phage genomes sequences, were first filtered by selecting the highest scoring match for each unique spacer sequence then merging any remaining redundant host-target matches, such that each spacer is represented only once in the dataset.

Classification of phages as nucleus-forming. Using the PCH45 shell (Gp202) and tubulin (Gp187) protein sequences as queries, we identified homologous proteins in GenBank *Caudovirales* genomes ≥150 kb via iterative hidden Markov model (HMM) searches using jackhmmer (from HMMER v.3.2.1)[53]. Phages encoding homologues of both shell and tubulin proteins (expect value < 10^{-10} for both) were classified as nucleus-forming. We then manually curated shell and tubulin protein alignments (multiple sequence alignment)[54] for only the shell and tubulin homologues occurring in nucleus-forming phages (Extended Data Fig. 4a). Using these alignments, we generated HMMs using HMMER3 v.3.2.1 (ref. [55]) and used the HMMs to classify target phage genomes and IMG/VR contigs as nucleus-forming (encoding matches to both shell and tubulin HMMs with expect values < 10^{-6}). Since many IMG/VR contigs represent incomplete phage genomes belonging to defined viral families (clusters), we also classified contigs (approximately 10% of the nucleus-forming matches) as belonging to nucleus-forming phages if there were other contigs within the viral cluster that encoded homologues of both shell and tubulin proteins.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available from the corresponding author on request. The genome sequence of bacteriophage PCH45 is available in GenBank under accession number MN334766. Data and R scripts for the bioinformatics analyses are available at https://github.com/JacksonLab/Jumbophages.

Received: 3 October 2019; Accepted: 17 October 2019;
Published online: 09 December 2019

References

1. Hille, F. et al. The biology of CRISPR-Cas: backward and forward. *Cell* **172**, 1239–1259 (2018).
2. Bryson, A. L. et al. Covalent modification of bacteriophage T4 DNA inhibits CRISPR-Cas9. *mBio* **6**, e00648 (2015).
3. Vlot, M. et al. Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR–Cas effector complexes. *Nucleic Acids Res.* **46**, 873–885 (2018).
4. Pawluk, A., Davidson, A. R. & Maxwell, K. L. Anti-CRISPR: discovery, mechanism and function. *Nat. Rev. Microbiol.* **16**, 12–17 (2018).
5. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013).
6. Chaikeeratisak, V. et al. Assembly of a nucleus-like structure during viral replication in bacteria. *Science* **355**, 194–197 (2017).
7. Chaikeeratisak, V. et al. The phage nucleus and tubulin spindle are conserved among large *Pseudomonas* phages. *Cell Rep.* **20**, 1563–1571 (2017).
8. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Classification and nomenclature of CRISPR-Cas systems: where from here? *CRISPR J.* **1**, 325–336 (2018).
9. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
10. Jackson, S. A. et al. CRISPR-Cas: adapting to change. *Science* **356**, eaal5056 (2017).
11. Brouns, S. J. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
12. Hale, C. R. et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945–956 (2009).
13. Garneau, J. E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
14. Patterson, A. G. et al. Quorum sensing controls adaptive immunity through the regulation of multiple CRISPR-Cas systems. *Mol. Cell* **64**, 1102–1108 (2016).
15. Yuan, Y. & Gao, M. Jumbo bacteriophages: an overview. *Front. Microbiol.* **8**, 403 (2017).
16. Mesyanzhinov, V. V. et al. The genome of bacteriophage φKZ of *Pseudomonas aeruginosa*. *J. Mol. Biol.* **317**, 1–19 (2002).
17. Ceyssens, P. J. et al. Development of giant bacteriophage φKZ is independent of the host transcription apparatus. *J. Virol.* **88**, 10501–10510 (2014).
18. Jackson, S. A., Birkholz, N., Malone, L. M. & Fineran, P. C. Imprecise spacer acquisition generates CRISPR-Cas immune diversity through primed adaptation. *Cell Host Microbe* **25**, 250–260 (2019).
19. Kraemer, J. A. et al. A phage tubulin assembles dynamic filaments by an atypical mechanism to center viral DNA within the host cell. *Cell* **149**, 1488–1499 (2012).
20. Erb, M. L. et al. A bacteriophage tubulin harnesses dynamic instability to center DNA in infected cells. *eLife* **3**, e03197 (2014).
21. Samai, P. et al. Co-transcriptional DNA and RNA cleavage during type III CRISPR-Cas immunity. *Cell* **161**, 1164–1174 (2015).
22. Goldberg, G. W., Jiang, W., Bikard, D. & Marraffini, L. A. Conditional tolerance of temperate phages via transcription-dependent CRISPR–Cas targeting. *Nature* **514**, 633–637 (2014).
23. Jia, N. et al. Type III-A CRISPR-Cas Csm complexes: assembly, periodic RNA cleavage, DNase activity regulation, and autoimmunity. *Mol. Cell* **73**, 264–277 (2019).
24. Liu, T. Y., Iavarone, A. T. & Doudna, J. A. RNA and DNA targeting by a reconstituted *Thermus thermophilus* type III-A CRISPR-Cas system. *PLoS ONE* **12**, e0170552 (2017).
25. Niewoehner, O. et al. Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548 (2017).
26. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* **357**, 605–609 (2017).
27. Rouillon, C., Athukoralage, J. S., Graham, S., Gruschow, S. & White, M. F. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *eLife* **7**, e36734 (2018).
28. Varble, A. & Marraffini, L. A. Three new Cs for CRISPR: collateral, communicate, cooperate. *Trends Genet.* **35**, 446–456 (2019).
29. Lau, R. K. et al. Structure and mechanism of a cyclic trinucleotide-activated bacterial endonuclease mediating bacteriophage immunity. Preprint at https://doi.org/10.1101/694703 (2019).
30. Silas, S. et al. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* **351**, aad4234 (2016).
31. Hynes, A. P., Villion, M. & Moineau, S. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. *Nat. Commun.* **5**, 4399 (2014).

32. Mendoza, S. D. et al. A nucleus-like compartment shields bacteriophage DNA from CRISPR-Cas and restriction nucleases. Preprint at https://doi.org/10.1101/370791 (2018).
33. Watson, B. N. J., Staals, R. H. J. & Fineran, P. C. CRISPR-Cas-mediated phage resistance enhances horizontal gene transfer by transduction. *mBio* **9**, e02406-17 (2018).
34. Frampton, R. A. et al. Identification of bacteriophages for biocontrol of the kiwifruit canker phytopathogen *Pseudomonas syringae* pv. actinidiae. *Appl. Environ. Microbiol.* **80**, 2216–2228 (2014).
35. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
36. Brettin, T. et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **5**, 8365 (2015).
37. Lowe, T. M. & Chan, P. P. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **44**, W54–W57 (2016).
38. Carver, T., Thomson, N., Bleasby, A., Berriman, M. & Parkhill, J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* **25**, 119–120 (2009).
39. Bao, Y., Chetvernin, V. & Tatusova, T. Improvements to pairwise sequence comparison (PASC): a genome-based web tool for virus classification. *Arch. Virol.* **159**, 3293–3304 (2014).
40. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).
41. Meier-Kolthoff, J. P. & Göker, M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics* **33**, 3396–3404 (2017).
42. Pawluk, A. et al. Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat. Microbiol.* **1**, 16085 (2016).
43. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
44. Hampton, H. G. et al. CRISPR-Cas gene-editing reveals RsmA and RsmC act through FlhDC to repress the SdhE flavinylation factor and control motility and prodigiosin production in *Serratia*. *Microbiology* **162**, 1047–1058 (2016).
45. Kaniga, K., Delor, I. & Cornelis, G. R. A wide-host-range suicide vector for improving reverse genetics in Gram-negative bacteria: inactivation of the *blaA* gene of *Yersinia enterocolitica*. *Gene* **109**, 137–141 (1991).
46. Watson, B. N. J. et al. Type I-F CRISPR-Cas resistance against virulent phage infection triggers abortive infection and provides population-level immunity. Preprint at https://doi.org/10.1101/679308 (2019).
47. Silva-Rocha, R. et al. The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* **41**, D666–D675 (2013).
48. Biswas, A., Staals, R. H., Morales, S. E., Fineran, P. C. & Brown, C. M. CRISPRDetect: a flexible algorithm to define CRISPR arrays. *BMC Genomics* **17**, 356 (2016).
49. Nicholson, T. J. et al. Bioinformatic evidence of widespread priming in type I and II CRISPR-Cas systems. *RNA Biol.* **16**, 566–576 (2019).
50. Paez-Espino, D. et al. IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Res.* **47**, D678–D686 (2019).
51. Rizk, G. & Lavenier, D. GASSST: global alignment short sequence search tool. *Bioinformatics* **26**, 2534–2540 (2010).
52. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
53. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).
54. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
55. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).

Acknowledgements

This work was supported by the Marsden Fund from the Royal Society of New Zealand and a University of Otago Research Grant. L.M.M. was supported by a University of Otago Doctoral Scholarship. We thank staff of the Otago Micro and Nano Imaging facility for assistance with electron and confocal microscopy and the Otago Genomics Facility for genome sequencing. We thank members of the Fineran laboratory for helpful discussions, S. Shehreen, T. Nicholson and X. Morgan for bioinformatics advice. We also acknowledge the use of the New Zealand eScience Infrastructure (NeSI) high-performance computing facilities in this research. NeSI's facilities are provided by and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation and Employment's Research Infrastructure programme.

Author contributions

L.M.M., S.L.W., C.W. and L.F.G. performed the experiments. L.M.M., S.L.W., S.A.J. and C.W. generated the strains and plasmids. L.M.M., S.L.W. and L.F.G. performed the microscopy. L.M.M. and S.A.J. performed the bioinformatics analysis with input from P.P.G. and P.C.F. L.M.M. and P.C.F. conceived the project with input from all authors. P.C.F. supervised the project. L.M.M. and P.C.F. wrote the manuscript. All authors edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-019-0612-5.

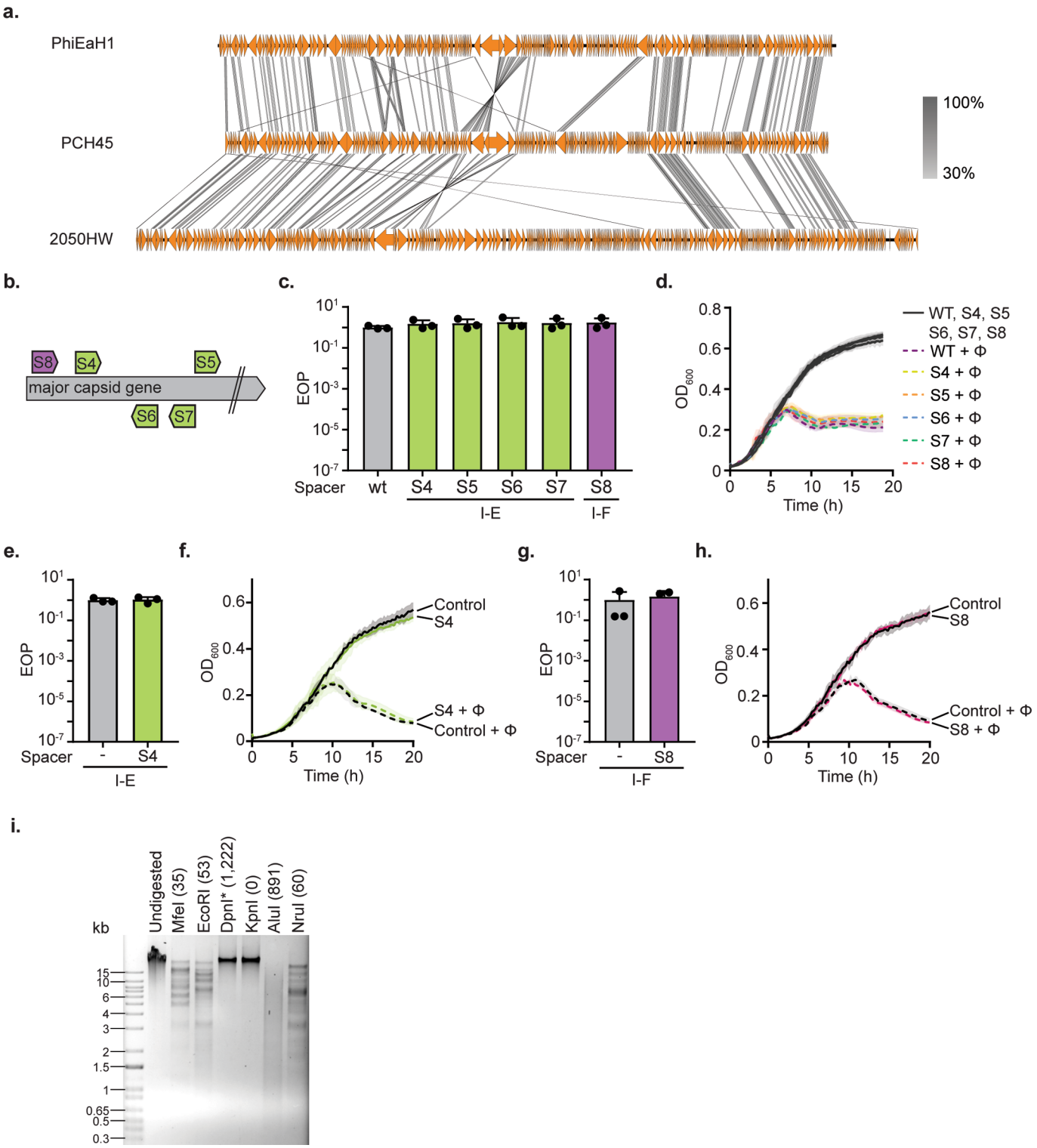
Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0612-5.

Correspondence and requests for materials should be addressed to P.C.F.

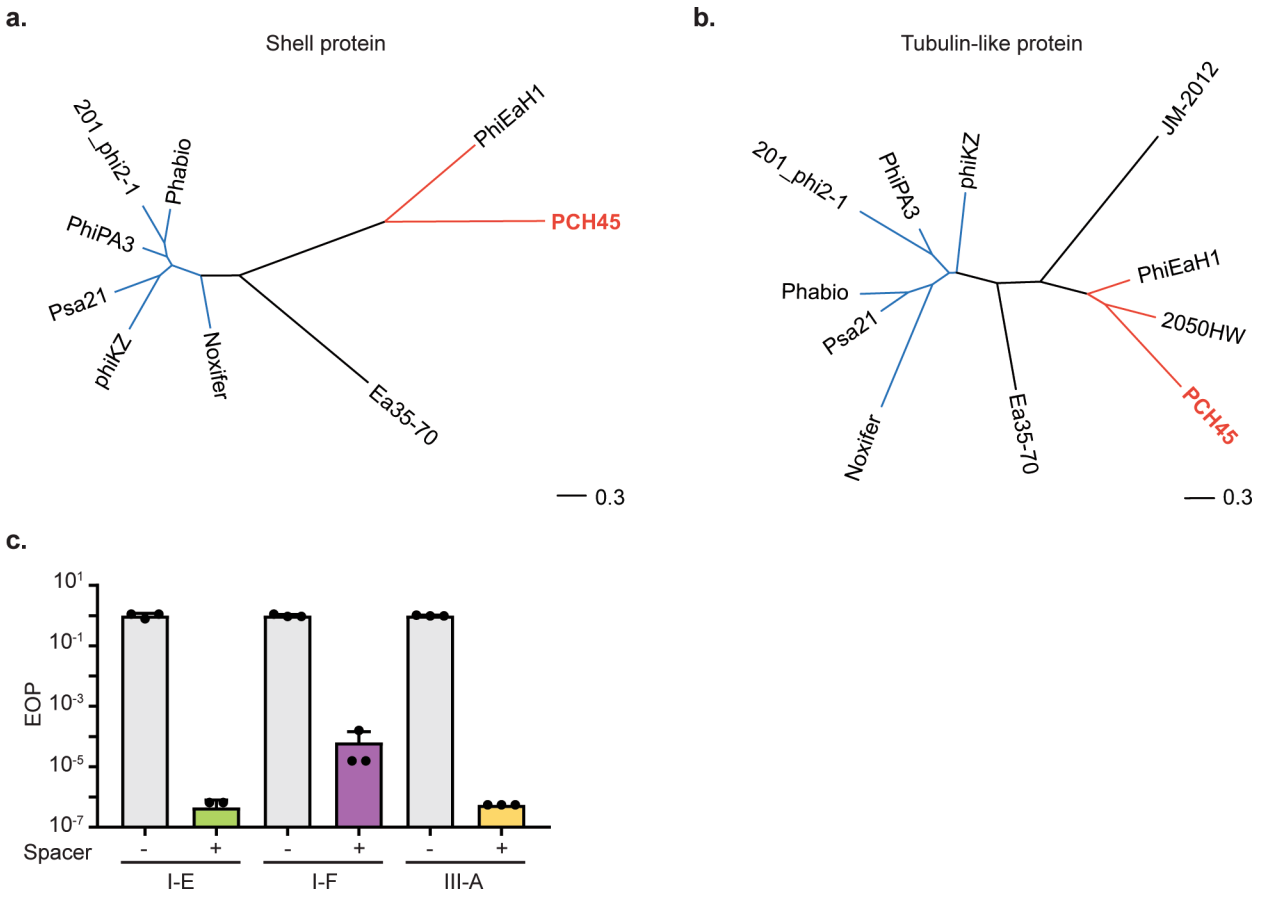
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

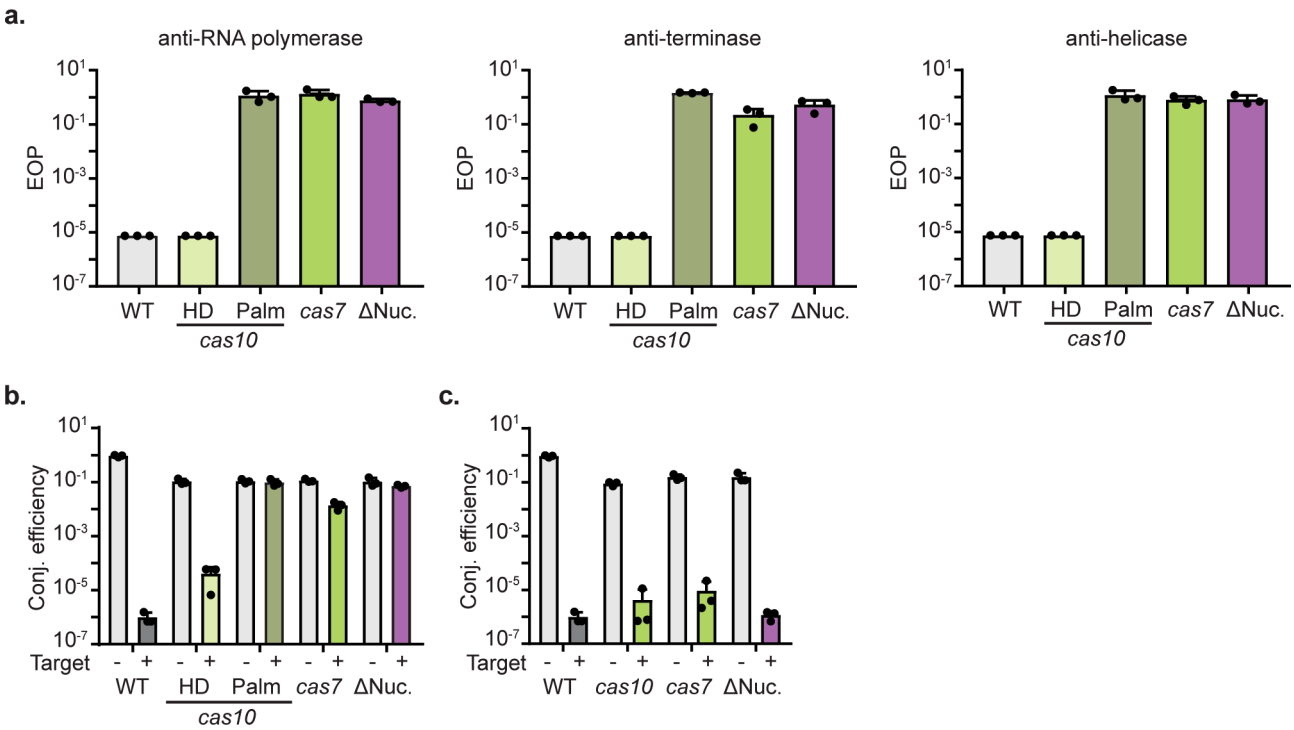
© The Author(s), under exclusive licence to Springer Nature Limited 2019



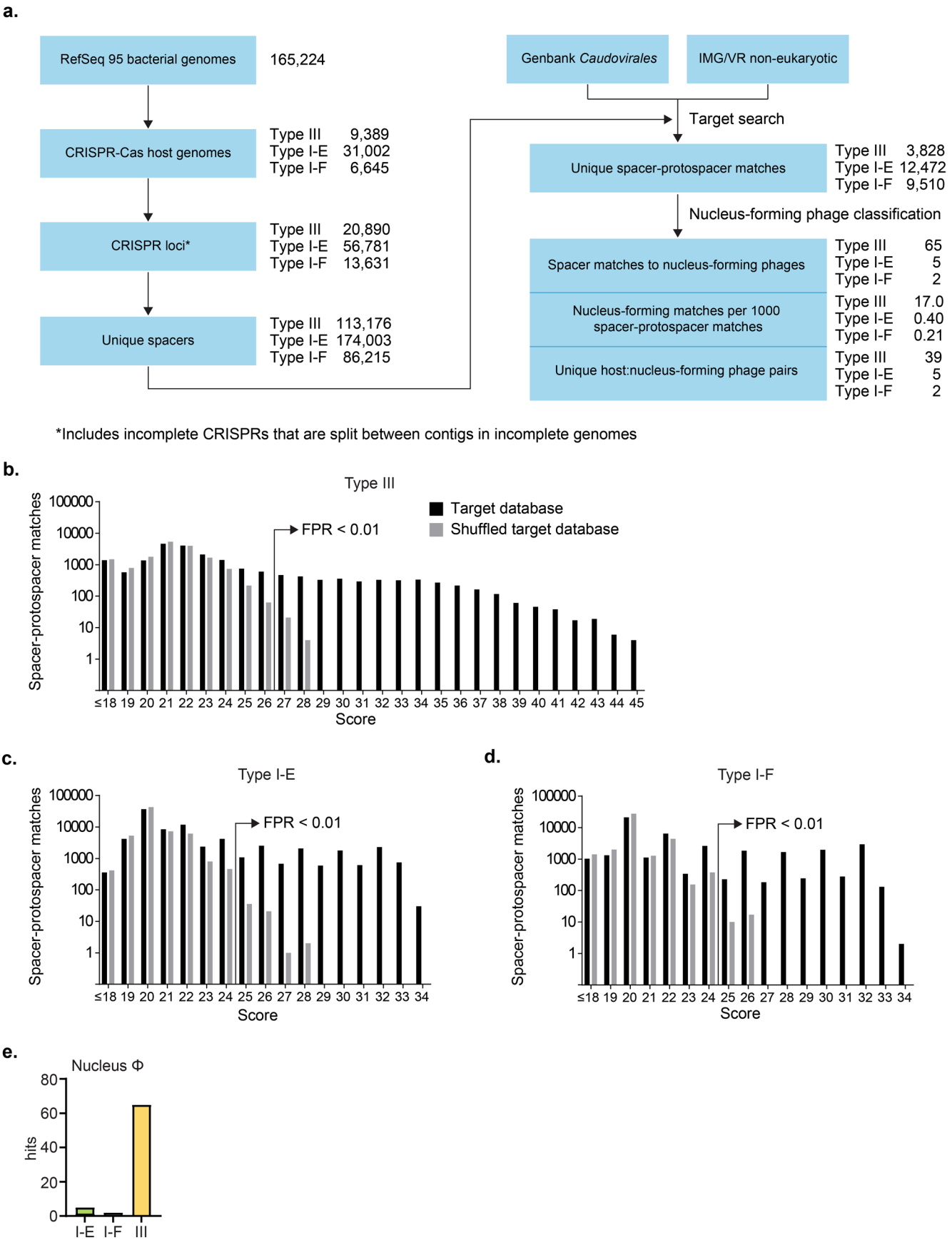
Extended Data Fig. 1 | The jumbo phage is resistant to type I CRISPR-Cas immunity. a. tblastx alignment of PCH45 with phages PhiEaH1 and 2050HW. **b**. Target location of chromosomally expressed anti-PCH45 type I-E (S4-7) and type I-F (S8) spacers targeting the major capsid gene (*gp033*). Phage resistance measured by **c**. EOP or **d**. plate reader assays for *Serratia* strains with type I-E, (S4, PCF591; S5, PCF593; S6, PCF545; S7, PCF544) and type I-F (S8, PCF548) infected with PCH45. Phage resistance measured by **e**. EOP or **f**. plate reader assays for *Serratia* carrying a type I-E (S4, pPF1460) spacer in mini-CRISPR arrays, infected with PCH45. Phage resistance measured by **g**. EOP or **h**. plate reader assays for *Serratia* carrying a type I-F (S8, pPF1461) spacer in a plasmid mini-CRISPR array, infected with PCH45. In c, e and g MOI=0.001. In c-h data are presented as mean ± s.d. (n=3 biologically independent samples). **i**. Restriction length fragment polymorphism (RLFP) analysis of phage gDNA treated with restriction enzymes MfeI, EcoRI, DpnI*, KpnI, AluI and NruI. Undigested PCH45 gDNA was used as a negative control. In parenthesis the number of restriction sites found in the genome of PCH45. (*): Cleaves only when the recognition motif is methylated. This experiment was performed three times with similar results and a representative gel is shown.



Extended Data Fig. 2 | The shell and tubulin proteins in *Serratia* jumbo phage PCH45 possess low sequence similarity to homologues encoded by other jumbo phages. Phylogenetic tree of **a**. the shell protein and **b**. PhuZ protein encoded by jumbo phages. The maximum likelihood trees were built for phage encoded shell and tubulin-like proteins (n=9 and n=11, respectively) using RaxML with 100 bootstrap replicates. The scale bar represents the approximate number of changes per amino acid position. **c**. EOP assay for *Serratia* strains PCF761 (*mCherry2-cas8e*), PCF763 (*mCherry2-cas8f)* and PCF765 (*mCherry2-cas10*) carrying type I-E, I-F and III-A anti-JS26 spacers in plasmids (pPF1485, pPF1489 and pPF1473 respectively). In c data are presented as mean ± s.d. (n=3 biologically independent samples).



Extended Data Fig. 3 | Type III RNA targeting provides protection against jumbo phage infection. a. EOP assay for type III-A mutant strains: *cas10*$^{H17A, N18A}$ (HD domain), *cas10*$^{D618A, D619A}$ (Palm domain), *cas7*^{D34A}, and the accessory nuclease knock out carrying an anti-PCH45 spacers (RNA polymerase beta subunit, S9; anti-terminase S10; and anti-helicase, S12) overexpressed *in trans* from a plasmid mini-CRISPR array. **b**. Conjugation efficiency assay (transconjugants/recipients) of plasmids pPF781 (untargeted control) and pPF1043 targeted by the type III-A CRISPR-Cas systems for *Serratia* strains. The type III-A mutants: PCF683 (*cas7*^{D34A}), PCF690 (*cas10* HD mutant), PCF691 (*cas10* Palm mutant) PCF686 (Δ accessory nuclease), and **c**. the chromosomal complementation with wild-type copies of the genes in PCF684 (*cas7*), PCF688 (*cas10*) and PCF687 (accessory nuclease). All data are presented as mean ± s.d. (n=3 biologically independent samples).



Extended Data Fig. 4 | See next page for caption

Extended Data Fig. 4 | Type III CRISPR arrays are enriched in jumbo phage-targeting spacers. a. Workflow used to obtain spacer-phage hits. Scores for spacer-target matches for targeted (black) and shuffled (grey) databases for **b**. type III **c**. type I-E and **d**. type I-F CRISPR-Cas systems. Scores with a false positive rate (FPR) < 0.01 were used as a cut-off to determine the spacer-protospacer hits; the FPR is defined as (the number of hits above the scoring threshold to the shuffled target database)/(the number of hits above the scoring threshold to the target database). **e**. Number of unique spacers in type I-E, I-F or type III systems matching nucleus-forming phages.

Strain Name	System	No. of new spacers	Sequence	Protospacer	Strand
PCF544	Type I-E	5	CTGTGAGCAGTGTTGAACTTCAGGTACTGAGT	Phage major capsid insert (*gp033_PCH45*)	-
			CTGTGACCGTCTCCGGGAGCTGCATGTGTCAG	plasmid backbone	-
			GCACAATTCTCATGTTTGACAGCTTATCATCG	plasmid backbone	-
			CAAATAAATTTTTTATGATTTCTCGAGCTCAT	plasmid backbone	-
			TCAGAGGTGGCGAAACCCGACAGGACTATAAA	pBR322origin	+
PCF545	Type I-E	4	GGACTCCTCCTTTATTTTATTTAGAATTCTGT	plasmid backbone	-
			CGCATGAACTCCTTGATGATGGCCATGTTATC	mCherry	-
			TTAGCTCACTCATTAGGCACAATTCTCATGTT	plasmid backbone	-
			TCCGTTTCCAGGATTTTGTGGTTCAGCAGCGC	Phage major capsid insert (*gp033_PCH45*)	-
PCF591	Type I-E	2	TTTCTGTAGAAGCTGCAACCATTCTGGGCCTG	Phage major capsid insert (*gp033_PCH45*)	+
			CTGTCAAACATGAGAATTGTGCCTAATGAGTG	plasmid backbone	+
PCF592	Type I-E	1	CCTTTCTCCAGATGGTCGATCGCGCTGTCAAA	Phage major capsid insert (*gp033_PCH45*)	-
PCF593	Type I-E	3	GGTGCACCGAACTTCACCGACTCTCTGGACCG	Phage major capsid insert (*gp033_PCH45*)	+
			TTGGACATCACCTCCCACAACGAGGACTACAC	mCherry	+
			CATTCTGCCGACATGGAAGCCATCACAAACGG	Undetermined target	
PCF547	Type I-F	1	ATTCTGGGCCTGGCTCACAATGGCGTCGGCAA	Phage major capsid insert (*gp033_PCH45*)	+
PCF548	Type I-F	2	GATAGCGGAACGGGAAGGCGACTGGAGTGCCA	*lacI*	-
			ATCACCGGCGGTCTTAACGTTCTGCCGCAGAA	Phage major capsid insert (*gp033_PCH45*)	+

Extended Data Fig. 5 | Spacer list in native CRISPR arrays.

Plasmid Name	Spacer	System	Spacer Sequence	Target
pPF1459	S1	Type I-E	CCTTTCTCCAGATGGTCGATCGCGCTGTCAAA	Major capsid gene (*gp033*)
pPF1460	S4	Type I-E	TTTCTGTAGAAGCTGCAACCATTCTGGGCCTG	Major capsid gene (*gp033*)
pPF1461	S8	Type I-F	ATCACCGGCGGTCTTAACGTTCTGCCGCAGAA	Major capsid gene (*gp033*)
pPF1462	S2	Type I-F	ATTCTGGGCCTGGCTCACAATGGCGTCGGCAA	Major capsid gene (*gp033*)
pPF1468	S12	Type III-A	CTTACGGATTTCTACTGGGACGCCATCGTACATG	DNA helicase (*gp217*)
pPF1466	S9	Type III-A	CGGATAGACACCATTAGGAACGGTATTGTCATGC	RNA polymerase beta subunit (*gp084*)
pPF1467	S3	Type III-A	GGTTTGCGCCCGGATGGAACTTCACGGTCATGTC	Major capsid gene (*gp033*)
pPF1469	S10	Type III-A	CCACGTACCGTTAACGATGATTTTACGACCACGG	Terminase large subunit (*gp159*)
pPF1470	S11	Type III-A	GTTTTTGTTACCGGCTTCGTTGTTGTCGTAGAAC	Tubulin-like gene (*gp187*)
pPF1473	S15	Type III-A	ATACAGTTTTTCGTAATAGTCCTCGTTGGCCTGC	DNA helicase (JS26)
pPF1485	S16	Type I-E	GCCGCTTTGTTGGCCTCCGCAAACTGGCGAAT	Tail length tape measure gene (JS26)
pPF1489	S17	Type I-F	GGTAATGGCACCCATGTCGCCCTGCAGCGCTT	Tail length tape measure gene (JS26)
pPF1994	S13	Type III-A	GACATGACCGTGAAGTTCCATCCGGGCGCAAACC	Reverse PS4 (Major capsid gene *gp033*)
pPF1995	S14	Type III-A	GACATGACCGTGAAGTTCCATCCGGGCGCAAACC	Reverse PS5 (Terminase large subunit *gp159*)
pPF1996	S11	Type III-A	GTTCTACGACAACAACGAAGCCGGTAACAAAAAC	Reverse PS6 (Tubulin-like gene *gp187*)

Extended Data Fig. 6 | Spacers expressed from mini-array in plasmid.

nature research

Corresponding author(s): Peter Fineran

Last updated by author(s): Oct 11, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The softwares employed in the bioinformatic analysis are detailed in the methods section and the script customized for this study is available on GitHub as detailed in the Data Availability statement in the manuscript.
Data analysis	The softwares employed in the bioinformatic analysis are detailed in the methods section and the script customized for this study is available on GitHub as detailed in the Data Availability statement in the manuscript.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The genome sequence of bacteriophage PCH45 is available in the GenBank database under accession number MN334766

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. All samples sizes are listed in the manuscript and are standard for microbiological analyses. Large differences were observed in most cases, so small samples sizes were sufficient.
Data exclusions	No data points were excluded from analyses.
Replication	Reproducibility of our experiments was determined by performing experiments with at least 3 biological replicates and we observed reproducible data.
Randomization	Any microbiological cultures subjected to different treatments (e.g. +/- phage infection) were randomly assigned to each group.
Blinding	Blinding was not required as no group allocation occurs

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	The study did not involve laboratory animals.
Wild animals	The study did not involve wild animals.
Field-collected samples	PCH45 Jumbophage was isolated from wastewater samples collected in Tahuna wastewater treatment plant (45°54'16.1"S; 170° 31'16.8"E, Dunedin, New Zealand).
Ethics oversight	No ethical approval was required

Note that full information on the approval of the study protocol must also be provided in the manuscript.