LETTER

doi:10.1038/nature11927

A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity

Kimberley D. Seed[1], David W. Lazinski[1], Stephen B. Calderwood[2,3] & Andrew Camilli[1]

Bacteriophages (or phages) are the most abundant biological entities on earth, and are estimated to outnumber their bacterial prey by tenfold[1]. The constant threat of phage predation has led to the evolution of a broad range of bacterial immunity mechanisms that in turn result in the evolution of diverse phage immune evasion strategies, leading to a dynamic co-evolutionary arms race[2,3]. Although bacterial innate immune mechanisms against phage abound, the only documented bacterial adaptive immune system is the CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated proteins) system, which provides sequence-specific protection from invading nucleic acids, including phage[4–11]. Here we show a remarkable turn of events, in which a phage-encoded CRISPR/Cas system is used to counteract a phage inhibitory chromosomal island of the bacterial host. A successful lytic infection by the phage is dependent on sequence identity between CRISPR spacers and the target chromosomal island. In the absence of such targeting, the phage-encoded CRISPR/Cas system can acquire new spacers to evolve rapidly and ensure effective targeting of the chromosomal island to restore phage replication.

Vibrio cholerae serogroup O1 is the primary causative agent of the severe diarrhoeal disease cholera, and lytic *V. cholerae* phages have been implicated in easing disease burden, particularly in the endemic region surrounding the Bay of Bengal[12,13]. We recently described the isolation of the ICP1 (for the International Centre for Diarrhoeal Disease Research, Bangladesh cholera phage 1)-related, *V. cholerae* O1-specific virulent myoviruses that are omnipresent among cholera patient rice-water stool samples collected at the ICDDR,B from 2001 to 2011 (ref. 14 and present study). *V. cholerae* readily evolves resistance to ICP1 predation through mutations in O1 antigen biosynthetic genes outside the human host; however, this mutational escape comes at a cost as virulence necessitates maintenance of the O1 antigen[15]. This dynamic between predation by ICP1 and virulence of *V. cholerae* O1, specifically in the context of human infection, provides a unique opportunity for discovery of novel bacterial immunity and phage immune evasion strategies. One bacterial defensive strategy against phages is the CRISPR/Cas system. CRISPR loci consist of an array of short direct repeats separated by highly variable spacer sequences of precise length corresponding to segments of previously captured foreign DNA (protospacers)[4,7,9]. CRISPR loci are found in ~40% and ~90% of sequenced bacterial and archaeal genomes, respectively[8,16]. The CRISPR array is transcribed and the transcript cleaved into small CRISPR RNAs (crRNAs) that, in conjunction with the Cas proteins, execute an efficient process of immunity in which foreign nucleic acids are recognized by hybridization to crRNAs and cleaved[4,7,8].

We isolated eleven ICP1-related phages from stools of cholera patients at the ICDDR,B (ref. 14 and present study), five of which encode a CRISPR/Cas system located between open reading frames (ORFs) 87 and 88 of the ancestral ICP1 genome[14]. The GC content of this CRISPR/Cas system is the same (~37%) as the rest of the ICP1 genome. The ICP1 CRISPR/Cas system consists of two CRISPR loci (designated CR1 and CR2) and six *cas* genes (Fig. 1a) whose organization and protein products are most homologous to Cas proteins of the type 1-F (*Yersinia pestis*) subtype system[17] (Supplementary Table 1). *V. cholerae* is divided into two biotypes, classical and El Tor, the former of which is associated with earlier pandemics and has since been replaced by the El Tor biotype[18]. The classical strain, *V. cholerae* O395, has a CRISPR/Cas system belonging to the type I-E (*Escherichia coli*) subtype[17], and to date there has not been any description of El Tor strains possessing a CRISPR/Cas system. Thus, the origin of the CRISPR/Cas system in ICP1 phage is unknown. Protospacer-adjacent motifs (PAMs) are type-specific, short conserved sequence motifs in the immediate vicinity of protospacers that are required for acquisition and targeting[7,9,11,19]. In contrast to the GG PAM reported for the type I-F CRISPR/Cas systems in bacteria[19], the protospacers targeted by the ICP1 CRISPR array have a GA PAM (Supplementary Fig. 1).

The majority of spacers in the ICP1 CRISPR show 100% identity to sequences within an 18-kilobase (kb) island found in a subset of *V. cholerae* strains that include the classical strain O395 isolated in India in 1964, El Tor strain MJ-1236 isolated in Bangladesh in 1994, and several El Tor strains collected at the ICDDR,B between 2001 and 2011 (Supplementary Table 2). The 18-kb island resembles the phage-inducible chromosomal islands (PICIs) of Gram-positive bacteria, including the prototype *Staphylococcus aureus* pathogenicity islands (SaPIs)[20,21]. SaPIs are induced to excise, circularize and replicate

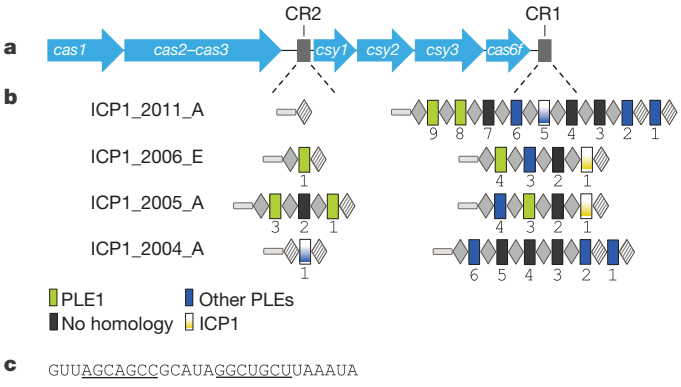

Figure 1 | Genomic organization of the ICP1 CRISPR/Cas system. a, The ICP1 phage CRISPR/Cas system consists of six *cas* genes and two CRISPR loci (CR1 and CR2). **b**, For each CRISPR locus, the repeat (28 bp) and spacer (32 bp) content is detailed as grey diamonds and coloured rectangles, respectively. Repeats (28 bp) that match the repeat consensus are shown in grey diamonds, and degenerate repeats are indicated in hatched grey diamonds. An AT-rich leader sequence precedes each CRISPR locus (grey rectangle). Spacers are coloured according to the percentage identity (solid represent 100% identity, gradient represents 81–97% identity). A fifth ICP1-related phage (ICP1_2003_A) has a genetically identical CRISPR/Cas system to ICP1_2004_A, and has been omitted for simplicity. **c**, The RNA sequence of the CR1 and CR2 consensus repeat with the partially palindromic sequence forming the predicted stem in the crRNA underlined.

[1]Howard Hughes Medical Institute and Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts 02111, USA. [2]Division of Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [3]Harvard Medical School, Boston, Massachusetts 02114, USA.

following infection by certain phages. They use varied mechanisms to interfere with the phage reproduction cycle to enable their own promiscuous spread[21], and this can protect the surrounding bacterial population from further phage predation. The organization of the *V. cholerae* 18-kb island targeted by the ICP1 CRISPR/Cas system is similar in length, base composition and organization to that observed in the SaPIs subset of PICIs, with an integrase homologue at one end and a GC content lower than that of the host species (37% compared to 47.5%). We therefore refer to the 18-kb element as the *V. cholerae* PICI-like element (PLE) (Fig. 2).

To address the functional relevance of the ICP1 CRISPR/Cas system, we focused on the interaction between the paired ICP1_2011_A phage and the *V. cholerae* O1 El Tor strain (harbouring PLE1) that were isolated from the same stool sample (for simplicity hereafter referred to as ICP1 and *V. cholerae* PLE$^+$). ICP1 has two CRISPR spacers (8 and 9) (Fig. 1b) that have 100% identity to sequences within the *V. cholerae* PLE (Fig. 2 and Supplementary Table 2). Using the standard soft agar overlay method, we found that ICP1 can plaque efficiently on *V. cholerae* PLE$^+$ (Fig. 3b). We used northern blot analysis to confirm that ICP1 crRNAs are transcribed and processed during *V. cholerae* infection (Supplementary Fig. 2). To test whether targeting of the PLE by the ICP1 CRISPR/Cas system affects phage fitness, we eliminated spacer 8 and 9 targeting. Spacer 8 targeting was disrupted by introducing silent mutations into its target within the PLE, generating *V. cholerae* PLE(8*) (Fig. 3a). We then infected this strain with a spontaneous ICP1 spacer 9 deletion mutant, referred to as ICP1(ΔS9). ICP1(ΔS9) was blocked for plaque formation on *V. cholerae* PLE(8*); however, it maintained wild-type plaquing efficiency on *V. cholerae* PLE$^+$ (Fig. 3b). Importantly, *V. cholerae* PLE(8*) is sensitive to plaque formation by ICP1 (Fig. 3b), which still harbours one spacer (S9) targeting the PLE. These results demonstrate that ICP1 CRISPR/Cas must target the PLE for destruction in order to effectively infect and form plaques, and that a single spacer that targets the PLE is sufficient to facilitate successful phage replication. A mutant in which PLE ORFs 7–20 were deleted was susceptible to infection by ICP1(ΔS9) with wild-type plaquing efficiency (Supplementary Fig. 3). This demonstrates that an intact PLE is required to inhibit ICP1 in the absence of CRISPR targeting. These results, in conjunction with the observation that PLE1 circularizes following ICP1 infection (Supplementary Fig. 4), further support our designation of the 18-kb island as a PICI-like element.

It has been well documented in the type I-E (*E. coli*) system that CRISPR interference requires an intact PAM and a fully complementary seed region (a non-contiguous 7 base pair (bp) sequence immediately adjacent to the PAM)[22]. To address the sequence requirements of the ICP1 CRISPR/Cas system we constructed a series of point mutations in the spacer 8 target in *V. cholerae* PLE that span the PAM, seed region and remainder of the target sequence, and determined their effect on immunity. In accordance with previous results, we found that single mutations within the PAM or the first four positions in the seed region immediately adjacent to the PAM abolish ICP1 CRISPR/Cas immunity (Supplementary Fig. 5). Interestingly, mutations of increasing distance from the PAM showed a concordant decreasing effect on immunity. Up to five mismatches outside of the seed region of the target are known to be tolerated in the type I-E system[22], and similarly we found that three and five mutations outside of the seed region were tolerated; however, eight mutations were not (Supplementary Fig. 5).

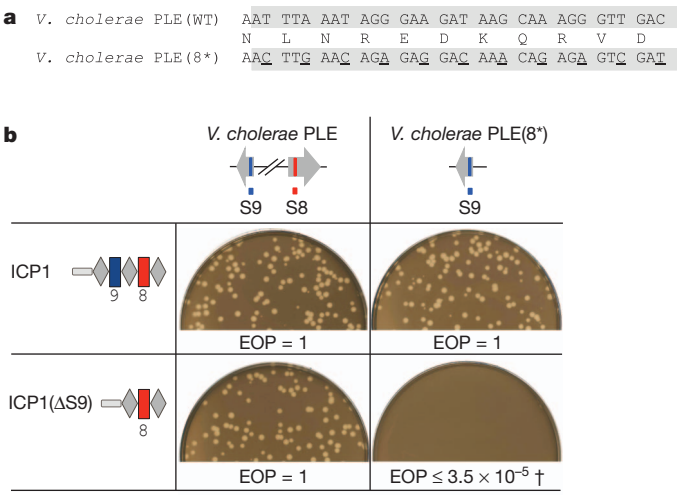

Figure 3 | Sequence-based targeting by the ICP1 CRISPR/Cas system is essential for lytic growth on *V. cholerae* PLE$^+$. **a**, Disruption of the *V. cholerae* PLE target protospacer generating *V. cholerae* PLE(8*). The 32 bp protospacer sequence is shaded in grey. **b**, The sensitivity of each strain (top row) to ICP1 or ICP1(ΔS9) (left column) is shown. Identity between the spacer and targeted protospacer is indicated by the red and blue rectangles. The efficiency of plaquing (EOP, which is the plaque count on the mutant host strain divided by that on the wild-type host strain) is indicated. A dagger indicates that the EOP is 10^{-5} or 10^{-8} depending on the presence of PLE in the host strain used for propagation as discussed in the text.

In experiments where the ICP1 CRISPR/Cas system could not target the *V. cholerae* PLE and therefore plaque formation was greatly reduced, we observed phage escape mutants at frequencies that were dependent on the host strain on which the phage had been previously propagated. When ICP1(ΔS9) was grown on a PLE$^+$ host before plaquing on *V. cholerae* PLE(8*), the efficiency of plaquing (EOP, which is the plaque count on the mutant host strain divided by that on the wild-type host strain) was 3.5×10^{-5}. The CRISPR loci from ten independent ICP1(ΔS9) escape mutants were sequenced, and in all cases, a new spacer was present at the leader end of the CRISPR CR1 array. Furthermore, the new spacers had 100% identity to sequences within the PLE (Fig. 2), and all newly integrated spacers target the PLE with the conserved GA dinucleotide PAM sequence (Supplementary Fig. 1b). The experimentally acquired spacers target both the coding and noncoding strands (Supplementary Table 3), although most (nine out of ten) target the coding strand. The pre-existing spacer (S8) (although mismatched in these experiments) also targets the coding strand; these data are in support of recent evidence that the DNA strand from which new protospacers are incorporated is heavily biased towards the existing protospacer orientation[23,24]. In contrast to when phage were propagated on a PLE$^+$ host before plaquing on *V. cholerae* PLE(8*), phage escape mutants were detected at a much lower frequency (EOP $= 1.1 \times 10^{-8}$) when ICP1(ΔS9) was grown on a *V. cholerae* PLE$^-$ host. This shows that new spacers targeting the PLE are incorporated into the CRISPR array during ICP1(ΔS9) infection of the PLE$^+$ host (the immunization process), and that an immune host possessing an untargeted PLE can subsequently be used to select for new ICP1 CRISPR acquisition events that confer targeting and thus restore phage replication. These results demonstrate that the ICP1

Figure 2 | Genomic organization of PLE1, a representative *V. cholerae* PLE targeted by the CRISPR/Cas system of ICP1-related phages. The integrase (*int*) is in blue, genes encoding hypothetical proteins (with numerical ORF designations) are grey. The locations of protospacers incorporated into the CRISPR locus as spacers 8 and 9 (S8 and S9 of ICP1_2011_A) are indicated in green above the map. The locations of experimentally acquired protospacers are shown below the map in red.

CRISPR/Cas system is fully functional as an adaptive immune evasion system that benefits the phage.

ICP1 has evolved to effectively target the *V. cholerae* PLE with an adaptive immune evasion system that has never before been shown to function in bacterial viruses. During ICP1 infection of *V. cholerae* PLE$^+$, PLE circularizes (Supplementary Fig. 4) and inhibits ICP1 through an unknown mechanism. To replicate successfully, ICP1 uses the CRISPR/Cas system to target the PLE for destruction. Because host cell death and DNA damage is inherent to lytic phage infection, CRISPR-mediated DNA cleavage of the PLE does not affect ICP1 infection. Sequencing data has been used to identify putative CRISPR arrays within a *Clostridium difficile* prophage[25], and more recently in metagenomic data sets of free viruses[26,27]. However, there is currently no evidence for expression or function of these putative arrays. We show that the ICP1-encoded CRISPR/Cas system actively and autonomously functions to inhibit host immunity and thereby permit lytic infection. This finding, in conjunction with the previous observations regarding the presence of CRISPR loci in other phages[25–27], suggests that the use of the so-called bacterial adaptive immune system by these bacterial predators may be an underappreciated immune evasion strategy in the unfolding phage versus host co-evolutionary arms race.

METHODS

Phages (ICP1_2011_A and ICP1_2006_E) and *V. cholerae* were isolated from cholera rice-water stool samples and propagated as described[14,15]. Genomic libraries were generated for phage and host strains as described[28] and sequenced using an Illumina HiSeq2000. A *V. cholerae* O1 El Tor isolate collected at the ICDDR,B in 2006, which was sequenced in this study and found to not harbour a PLE, was used as the PLE$^-$ host for propagation experiments. We used the CRISPRFinder program[16] to identify CRISPR loci. WebLogo[29] was used to generate sequence logos for identification of the PAM. Point mutations were constructed using splicing by overlap extension (SOE) PCR and introduced using pCVD442-lac as previously described[15]. The PLE1 deletion construct (missing 8.6 kb including ORFs 7–20) was constructed using SOE PCR and introduced by natural transformation with subsequent deletion of the antibiotic-resistance marker using the FLP recombinase method as described[30]. ICP1(ΔS9) was identified by screening for alterations in the CRISPR array by PCR following growth on *V. cholerae* PLE$^+$. RNA was purified using the mirVana kit (Ambion) at the indicated times and run on 12% polyacrylamide urea gels. Northern blots were pre-hybridized in Ultrahyb-oligo (Ambion) and hybridization was carried out at 37 °C overnight using 32-nucleotide 5′ end-labelled DNA probes (generated with [γ-32P]ATP and T4 polynucleotide kinase) complementary to spacers 8 and 6.

Received 9 November 2012; accepted 23 January 2013.

1. Brüssow, H. & Hendrix, R. W. Phage genomics: small is beautiful. *Cell* **108,** 13–16 (2002).
2. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Rev. Microbiol.* **8,** 317–327 (2010).
3. Stern, A. & Sorek, R. The phage-host arms race: shaping the evolution of microbes. *Bioessays* **33,** 43–51 (2011).
4. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327,** 167–170 (2010).
5. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
6. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322,** 1843–1845 (2008).
7. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* **45,** 273–297 (2011).
8. van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M. & Brouns, S. J. J. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* **34,** 401–407 (2009).
9. Karginov, F. V. & Hannon, G. J. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell* **37,** 7–19 (2010).
10. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).
11. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Rev. Genet.* **11,** 181–190 (2010).
12. Faruque, S. M. *et al.* Self-limiting nature of seasonal cholera epidemics: role of host-mediated amplification of phage. *Proc. Natl Acad. Sci. USA* **102,** 6119–6124 (2005).
13. Nelson, E. J., Harris, J. B., Morris, J. G., Calderwood, S. B. & Camilli, A. Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nature Rev. Microbiol.* **7,** 693–702 (2009).
14. Seed, K. D. *et al.* Evidence of a dominant lineage of *Vibrio cholerae*-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. *MBio* **2,** e00334–10 (2011).
15. Seed, K. D. *et al.* Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in *Vibrio cholerae* O1. *PLoS Pathog.* **8,** e1002917 (2012).
16. Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8,** 172 (2007).
17. Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nature Rev. Microbiol.* **9,** 467–477 (2011).
18. Longini, I. M. *et al.* Epidemic and endemic cholera trends over a 33-year period in Bangladesh. *J. Infect. Dis.* **186,** 246–251 (2002).
19. Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–740 (2009).
20. Novick, R. P., Christie, G. E. & Penadés, J. R. The phage-related chromosomal islands of Gram-positive bacteria. *Nature Rev. Microbiol.* **8,** 541–551 (2010).
21. Ram, G. *et al.* Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc. Natl Acad. Sci. USA* **109,** 16300–16305 (2012).
22. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA* **108,** 10098–10103 (2011).
23. Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Commun.* **3,** 945–947 (2012).
24. Swarts, D. C., Mosterd, C., van Passel, M. W. J. & Brouns, S. J. J. CRISPR interference directs strand specific spacer acquisition. *PLoS ONE* **7,** e35888 (2012).
25. Sebaihia, M. *et al.* The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature Genet.* **38,** 779–786 (2006).
26. Minot, S. *et al.* The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* **21,** 1616–1625 (2011).
27. Garcia-Heredia, I. *et al.* Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. *PLoS ONE* **7,** e33802 (2012).
28. Lazinski, D. W. L. & Camilli, A. Homopolymer tail-mediated ligation PCR: a streamlined and highly efficient method for DNA cloning and library construction. *Biotechniques* **54,** 25–34 (2013).
29. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).
30. De Souza Silva, O. & Blokesch, M. Genetic manipulation of *Vibrio cholerae* by combining natural transformation with FLP recombination. *Plasmid* **64,** 186–195 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank the Tufts University Core Facility for sequencing and computational support. This work was supported by US National Institutes of Health grants AI055058 (A.C.), AI045746 (A.C.) and AI058935 (S.B.C.). A.C. is a Howard Hughes Medical Institute Investigator.

Author Contributions K.D.S. and D.W.L. performed experiments. K.D.S., D.W.L. and A.C. designed experiments. K.D.S. and A.C. wrote the manuscript. S.B.C. provided materials. All authors discussed the results and commented on the manuscript.

Author Information The sequences for the CRISPR/Cas system in ICP1_2011_A and ICP1_2006_E have been deposited at GenBank (accession numbers KC152959 and KC152958, respectively). The sequences for the *V. cholerae* PLEs identified in clinical isolates from the ICDDR,B have been deposited at GenBank/EMBL/DDBJ under the accession numbers KC152960 (PLE1) and KC152961 (PLE2). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.C. (andrew.camilli@tufts.edu).