Title
Phage Wars: the molecular interactions underlying the arms race between a lytic bacteriophage and epidemic Vibrio cholerae

Author
McKitterick, Amelia Chase

Publication Date
2019

Peer reviewed|Thesis/dissertation

Phage Wars: the molecular interactions underlying the arms race between a lytic bacteriophage and epidemic *Vibrio cholerae*

By

Amelia Chase McKitterick

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

In

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Kimberley Seed, Chair
Professor Jeffery Cox
Professor Britt Glaunsinger
Professor Britt Koskella

Spring 2019

Abstract

Phage Wars: the molecular interactions underlying the arms race between a lytic bacteriophage and epidemic *Vibrio cholerae*

By

Amelia McKitterick

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Kimberley Seed, Chair

Viruses infect all living organisms, and in response, hosts have had to adapt to the constant threat of viral infection. The interactions between viruses and their hosts shape the evolutionary trajectory of both entities over time. Phages, which are viruses that infect bacteria, are the most prevalent infectious agents and can vastly outnumber their bacterial hosts in a given environment. One such target of phage predation is the pathogen *Vibrio cholerae*, which can be co-isolated with phages both from cholera patient stool samples, as well as from aquatic reservoirs in regions where cholera is endemic. Analysis of the phages that are shed in cholera patient stools indicates that one phage, ICP1, is the most dominant phage that infects epidemic *V. cholerae*, at least in Bangladesh. In response to the constant threat of infection, epidemic *V. cholerae* has acquired an anti-phage island, referred to as PLE, that specifically blocks ICP1 infection. Here, we use the interactions between ICP1 and *V. cholerae* PLE as a model system to examine the molecular interactions between a virulent phage and its bacterial host. Using natural isolates enables us to comprehend the ways in which this molecular arms race has influenced the evolution of both ICP1 and epidemic *V. cholerae*. In the first chapter, we identify the molecular specificity of the PLE response to ICP1 during infection, demonstrating that this anti-phage island is highly evolved to activate only in response to the dominant phage ICP1. In the second chapter, we identify additional ICP1-PLE molecular interactions that facilitate PLE escape from ICP1-mediated host takeover processes. In the third chapter, we examine the ICP1-encoded CRISPR-Cas adaptive immune system that is specifically deployed to processively degrade PLE and allow for ICP1 to productively infect the PLE (+) *V. cholerae* host. Overall, the work presented here updates our understanding of the mechanisms by which anti-phage islands, such as PLE, have the capacity to exploit their predatory phage to mobilize and influence the evolution of pathogenic populations.

Table of Contents

Chapter 1

The host-virus conflict

For life to be successful, cells have to be able to successfully replicate and divide. Although there is much debate as to whether or not viruses can be considered alive, many of the same principles of survival and adaptation apply to them. Indeed, the intersection between host and virus plays a large role in the evolution of all organisms as well as their infectious parasites, both in terms of acquisition of mechanisms to inhibit successful viral infections as well as mechanisms by which viruses can avoid host defenses. For most viruses, the central tenants of the lifecycle include introduction of the genome, expression of virally encoded genes, replication of the genome, and packaging of newly synthesized genomes into viral particles that are then released back into the environment.

The most common viruses are bacteriophages, or phages, which are viruses that infect bacterial hosts and can reach titers of over 10^{30} in some environments[1]. Phages can be broadly grouped by two different lifecycles: lytic phages and temperate phages. Lytic phages inject their genomes into their bacterial host, undergo the replication program, and lyse their host cells when the new phage genomes are packaged to enable further spread. While temperate phages have the ability to enter into the lytic cycle, they also have the ability to stably integrate into the bacterial host chromosome and lysogenize, which facilitates vertical spread of the phage through the bacterial population. Upon stimulation, these lysogens can then exit the bacterial chromosome and enter into the lytic cycle, replicate, and lyse the host.

The ability of phages to integrate into bacterial chromosomes has broad implications for the fitness of their microbial hosts. In order to be favorable to their hosts, many temperate phages encode genes that boost the fitness of the bacterial host, such as virulence genes[2]. Additionally, being integrated into host chromosomes also makes temperate phages susceptible to host-mediated accumulation of mutations, such as those that arise at low frequencies during bacterial replication, or recombination[3]. Degenerate temperate phages often can be identified as losing core components of the phage machinery that would allow for activation of a lytic cycle and propagation.

Another interesting development for temperate phages is domestication by bacterial hosts to be used to further defend against infecting phages[4]. Although not directly proven, many phage defense islands are hypothesized to have arisen from temperate phages due to conserved aspects of their lifecycles. Phage inducible chromosomal islands (PICIs) are prototypical phage defense islands, in that they are activated by the lytic replication of a helper phage, and enter into an excision, replication, and packaging cycle[5].

The prototypical PICI, *Staphylococcus aureus* pathogenicity islands (SaPIs), are genomic islands that, in the absence of helper phage, are transcriptionally repressed by a SaPI encoded master repressor that binds to a site of divergent transcription[6]. This repressor is anti-repressed by the physical interaction with a helper phage produced protein, thus enabling SaPI activation only in the presence of a replicating helper phage[7]. A wide swath of *S. aureus* phages encode proteins that SaPIs have evolved to recognize as an inducing cue, increasing the ability of one island to respond to infection by a myriad of sources[8]. Once de-repressed, the SaPI expresses its SaPI-encoded integrase and excisionase, which work together to direct excision of the island out of the bacterial chromosome into a circular episome[9]. This excised island then makes use of the replication machinery encoded by the bacterial host to replicate[6]. Concurrently, the

infecting helper phage is also progressing through its own lifecycle and producing phage gene products, such as structural proteins and packaging machinery. These structural proteins are then hijacked by the SaPI to package the replicated SaPI genomes and to horizontally transduce the packaged particles to neighboring cells[10–12]. While SaPIs do not fully inhibit the ability of the helper phage to complete its lifecycle, they do diminish the success of these phages[13].

As mobilizable islands SaPIs have central roles in transmission of virulence-associated genes. In addition to these pathogenicity islands that hijack phage machinery to further their spread, phages overall increase the capacity for bacterial horizontal gene transfer (HGT)[14]. Long harnessed as a molecular cloning technique, phages have been demonstrated to transfer DNA from one host to another through both generalized and specialized transduction in which regions of the bacterial host chromosome are incidentally packaged into phage capsids. Independent of phage packaging, plasmids released from lysed bacterial cells have also been demonstrated to passively transfer from one host to another following infection of the original host[15].

Despite the benefits of HGT on bacterial communities, phage blooms within bacterial populations can have devastating effects on the total diversity of the community[16]. Shifts in community members have been demonstrated to have drastic effects, such as what is seen when phages disrupt gut microbiome communities, leading to negative health outcomes for the human hosts[17].

Altogether, due to the dramatic effects that phages can have on their bacterial hosts, phage defense mechanisms are imperative to provide protection to both the individual host but also to the population, especially for bacteria that are under constant threat by phage predation. *Vibrio cholerae*, the causative agent of the diarrheal disease cholera, is one such pathogen that is consistently surrounded by phages, both in the aquatic reservoir where it presides, as well as in the human gut during disease[18,19]. Cholera is transmitted via the fecal-oral route: following ingestion of contaminated water *V. cholerae* colonizes the small intestine and produces cholera toxin, a temperate phage-encoded virulence factor, which induces profuse, watery diarrhea in the human host, ultimately returning *V. cholerae* back to the environment or facilitating person-to-person spread[20]. Cholera epidemics are often proceeded by environmental disasters and population displacement, but cholera is still endemic many regions in the world, such as Bangladesh, where outbreaks occur biannually.

Analysis of the phages that are shed in cholera patient stool samples from a diarrheal hospital in Dhaka, Bangladesh has revealed that the dominant phage that preys on *V. cholerae* is ICP1[19]. ICP1 is a lytic myovirus that is repeatedly co-isolated with *V. cholerae* from patient stools. Isolates recovered from cholera patient stool samples over a 16 year period has revealed that the ICP1 genome is remarkably highly conserved[21]. Overall, a typical ICP1 isolate is predicted to encode approximately 230 gene products across 125 kb, of which 185 are shared among all isolates. Remarkably, only around 17% of all ICP1 gene products have a predicted structure or function, making analysis of individual isolates challenging[21].

ICP1 is able to form plaques on epidemic *V. cholerae*, as the phage receptor is the O1 antigen, a requirement for *V. cholerae* virulence[19]. In a single round of infection, ICP1 produces on average 90 infectious virions over the course of 25 minutes of infection[22]. ICP1 is predicted to undergo two styles of replication: origin dependent

replication by 8 minutes post infection, followed by a switch to rolling circle replication by 16 minutes post infection [23]. Ultimately, over the course the infection, ICP1 has the capacity to lyse a *V. cholerae* culture and decimate the population.

Remarkably, ICP1 is the only demonstrated phage that encodes its own CRISPR-Cas system[24,25]. Briefly, CIRPSR-Cas systems act as an adaptive immune system that targets and cleaves invading nucleotides in a sequence-dependent manner[26], thus spacers that are incorporated into CRISPR arrays act as molecular memories that define the targeting capacity of the array[27,28]. The ICP1-encoded Type 1-F system is expressed *de novo* upon phage infection and has the capacity to acquire new spacers[24]. While not every ICP1 isolate encodes a CRISPR-Cas system, analysis of the ICP1 CRISPR arrays from clinical cholera patient stool samples indicates that the majority of spacers target a genomic island found exclusively in epidemic *V. cholerae*[24]*.* Functional analysis of this genomic island revealed that it is a specific anti-phage island that defends *V. cholerae* from ICP1 infection referred to as the phage inducible chromosomal island like element (PLE)[22,24].

Broadly speaking, in the absence of CRISPR, PLE blocks plaque formation by ICP1 but does not provide protection against unrelated cholera phages, suggesting that this island has evolved to only defend against the most prevalent predatory phage that targets *V. cholerae*[22]. Aside from the binary result of the ability of ICP1 to form a plaque on a PLE (+) *V. cholerae* lawn, there are many differences between the course of ICP1 infection of an individual PLE (+) and PLE (-) cell. Upon engaging with the host PLE (+) *V. cholerae*, ICP1 still injects its genome and begins to express phage encoded genes. Some aspect of the ICP1 program induces PLE into activity: first, PLE excises from the *V. cholerae* chromosome and circularizes, which can be detected within five minutes of ICP1 infection. Once excised, PLE begins to replicate to high copy. Interestingly, the ability of PLE to replicate is inversely correlated with the ability of ICP1 to replicate, thus diminishing the ability of ICP1 to successfully produce infectious virions[23]. Ultimately, PLE is able to inhibit ICP1 and also induce an accelerated, culture-wide lysis of infected cells within 20 minutes of ICP1 infection, so the infected cell still dies but no ICP1 progeny are produced, thus protecting the overall *V. cholerae* population and inhibiting plaque formation[22]. Application of lysate from antibiotic marked PLE (+) *V. cholerae* to naïve *V. cholerae* has also demonstrated the HGT of the marked PLE to a new host in an O1 antigen-dependent manner, suggesting that PLE also steals structural components from ICP1 to facilitate transduction. While ICP1-mediated CRISPR targeting of PLE interferes with PLE and allows for ICP1 plaque formation on PLE (+) *V. cholerae*, PLE is still able to transduce, hinting that there is more to be elucidated about CRISPR activity against PLE.

Five genetically distinct PLEs have been identified and characterized by their shared genetic content, lifecycle, and ability to block ICP1[22]. These five PLEs have a unique temporal distribution, where each PLE successively appears and is replaced within epidemic populations, suggesting that they evolve over time in order to combat ICP1. In addition to the multi-decade record of PLEs, ICP1 isolates over time additionally demonstrate modularity in the accessory genome[21]. As ICP1 and PLE (+) *V. cholerae* are consistently co-isolated from the same cholera patient stool samples, along with the long genetic history of natural isolates, the interaction between the two makes for an elegant model system. Both ICP1 and PLE are genetically manipulatable,

allowing for exploration of both the intrinsic biology of each individual part as well as the evolutionary pressures that each entity exerts on the other, driving adaptation in both phage and epidemic *V. cholerae* populations.

Here, using the interactions between *V. cholerae* PLE and ICP1, we discover and characterize the molecular interactions between PLE and ICP1 that underlie this molecular arms race. In Chapter 1, we examine the molecular markers of the specificity between ICP1 and PLE by identifying a phage-encoded protein that is specific to ICP1 that PLE hijacks to direct PLE excision during phage infection. In Chapter 2, we reveal the ways in which PLE is able to overcome host takeover by ICP1. We identify a phage-encoded gene that is necessary for PLE replication and reveal that mobilization of PLE through excision or replication is necessary to avoid ICP1-encoded nucleases. In Chapter 3, we use the interaction between ICP1 and PLE to gain a deeper understanding of the natural interference ability of the phage-encoded CRISPR-Cas system, demonstrating the ability of the phage-encoded Cas2-3 helicase/nuclease to processively degrade DNA *in vivo*. The overall work presented here contributes to our understanding of the conflict between viruses and their hosts, including the evolution of phage defense, the ways in which mobile genetic elements and phage defense islands can exploit their helper phages, and the biological function of endogenous CRISPR-Cas systems.

Chapter 2

Anti-phage islands force their target phage to directly mediate island excision and spread[29]

Amelia C. McKitterick and Kimberley D. Seed

Summary

Vibrio cholerae, the causative agent of the diarrheal disease cholera, is antagonized by the lytic phage ICP1 in the aquatic environment and in human hosts. Mobile genetic elements called PLEs (phage-inducible chromosomal island-like elements) protect *V. cholerae* from ICP1 infection and initiate their anti-phage response by excising from the chromosome. Here, we show that PLE 1 encodes a large serine recombinase, Int, that exploits an ICP1-specific protein as a recombination directionality factor (RDF) to excise in response to phage infection. We show that this phage-encoded protein is sufficient to direct Int-mediated recombination *in vitro* and that it is highly conserved in all sequenced ICP1 genomes. Our results uncover an aspect of the molecular specificity underlying the conflict between a single predatory phage and *V. cholerae* PLE and contribute to our understanding of long-term evolution between phage and their bacterial hosts in nature.

Introduction

Introduction

Bacteria and phage are in a constant battle for survival that shapes the evolution of both populations over time[30–32]. Bacteria evolve to overcome the challenges of phage predation through a variety of mechanisms, including phage receptor variability, DNA degradation mechanisms (e.g. restriction-modification systems and CRISPR-Cas systems), and phage inducible chromosomal islands (PICIs)[5,33]. While the former programs are deployed before phage infection to prevent the phage from completing its lifecycle, PICIs respond to a specific cue expressed by a target phage upon infection. PICIs are typically thought of as phage parasites that provide a fitness advantage to their host bacterium by limiting phage proliferation[34] and by carrying important virulence genes, such as those found in *Staphylococcus aureus*[13]. Upon induction, PICIs excise from the host chromosome, replicate, and redirect the target phage's packaging machinery, ultimately inhibiting phage replication and enabling PICI transduction[6,7,11,12,35].

Vibrio cholerae, the causative agent of the diarrheal disease cholera, interacts with predatory phages in the aquatic environment as well as in the human host[20,36], leading to speculation that phages influence cholera epidemic dynamics[18,37]. In particular, the lytic phage ICP1 has been recovered from cholera patient stool and water samples over at least 12 years in Bangladesh[19,24,25] and is consequently considered a persistent predator of epidemic *V. cholerae* in this region. In response to persistent ICP1 predation, *V. cholerae* has acquired the phage-inducible chromosomal island-like element (PLE)[24]. The five PLEs identified in *V. cholerae* isolates spanning a >60-year collection period exhibit a common, ICP1-dependent response, which initiates with the integrated PLE excising from the chromosome and circularizing[22] (Fig. 1.1a). PLE excision facilitates mobilization as PLEs are transduced following ICP1 infection, permitting their spread to *V. cholerae* recipients[22]. Although the underlying mechanisms are not known, PLEs block the phage replication program in what appears to be an ICP1-specific manner, as other tested cholera phages do not stimulate PLE circularization and are not blocked by PLEs[22]. Further, protection against ICP1 is absolute—PLE$^+$ cells produce no progeny phage. The transmission costs imposed by PLEs are a significant burden to ICP1 in nature, as some ICP1 isolates have acquired a CRISPR-Cas system to target PLEs, which allows ICP1 to persist in spite of PLE[18] .

Here, due to the apparent specificity and conservation of PLE circularization in response to ICP1, we set out to characterize the mechanism governing this response. We attribute PLE mobility to a PLE-encoded large serine recombinase, Int, that hijacks a specific phage protein, which we refer to as PexA, to direct PLE excision during ICP1 infection. We validate the functionality and specificity of this unique recombination system, in which the recombinase and recombination directionality factor (RDF) are encoded in separate genomes. Additionally, we show that PexA is also hijacked to trigger the excision of PLEs found in *V. cholerae* isolates recovered decades ago, highlighting the continued arms race that shapes long-term evolutionary trajectories of ICP1 and epidemic *V. cholerae*.

Results

PLE-encoded Int is necessary for PLE circularization

PLE encodes a gene product (Int) with an N-terminal serine recombinase domain and a large C-terminus containing a putative zinc ribbon domain and coiled-coiled motif characteristic of large serine recombinases (LSRs)[38]. Typically found in temperate phages, LSRs have the ability to catalyze recombination between attachment (*att*) sites[39–41]. Only the LSR is required to catalyze recombination between episomal (*attP*) and chromosomal (*attC*) sites, leading to integration of the episome into the host chromosome. To reverse this process, a recombination directionality factor (RDF) is required to physically interact with the LSR and direct the LSR to recombine the left

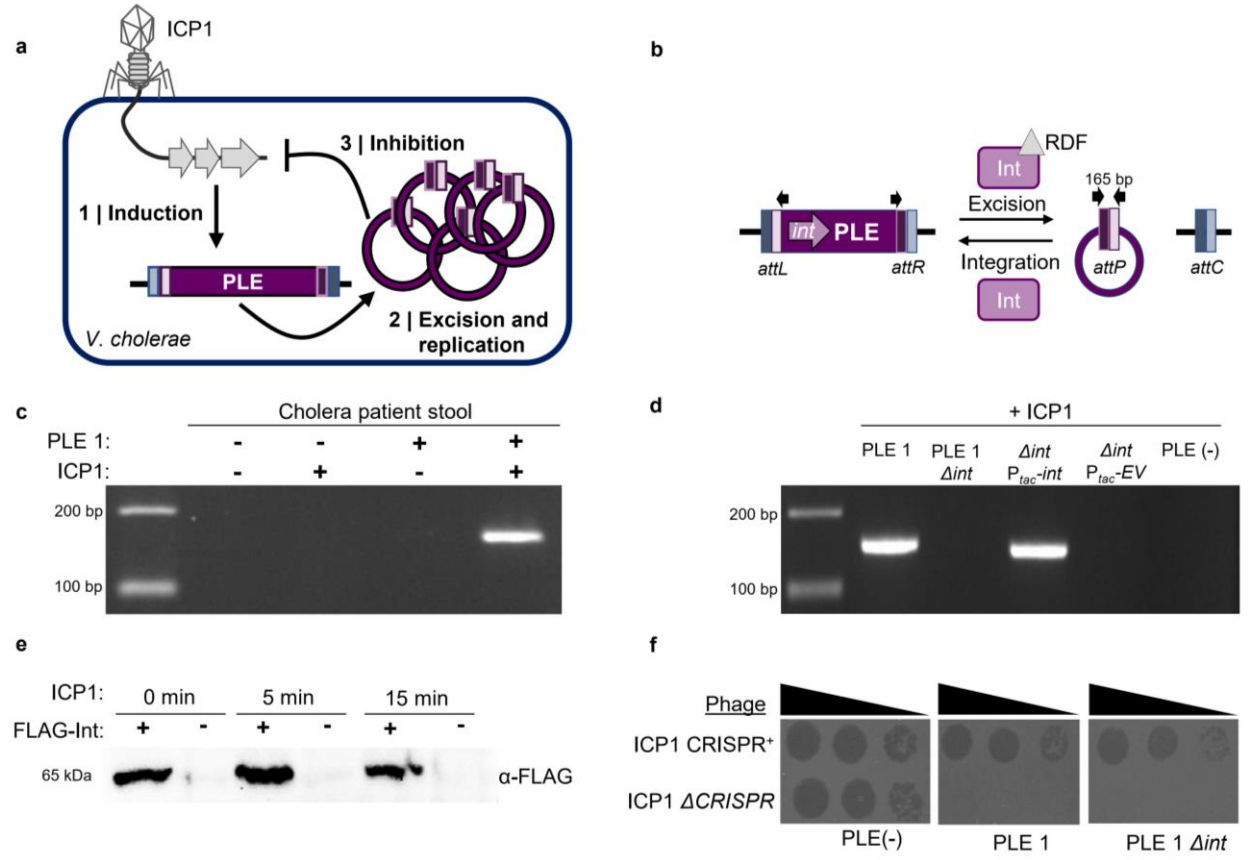


Figure 1.1. PLE circularizes during ICP1 infection and this response is dependent on PLE-encoded Int. a, ICP1 injects its DNA into a PLE$^+$ *V. cholerae* cell, leading to (1) PLE induction. (2) Induced PLE excises from the chromosome and circularizes, replicating to high copy number[22]. Through an unknown mechanism, (3) PLE inhibits the ICP1 replication program and protects the *V. cholerae* population from ICP1 predation. **b**, Model of LSR/RDF mediated integration and excision of PLE and the resulting *att* sites. Black arrows indicate primers used to detect PLE circularization. **c**, Circularized PLE, which is detected by PCR using outward-facing primers internal to the PLE, is found in cholera patient stool samples when PLE 1 and ICP1 are both present. **d**, Detection of PLE 1 circularization during ICP1 infection. *V. cholerae* with the PLE 1 variant indicated was infected with ICP1 at a multiplicity of infection (MOI) of 5 and harvested 5 minutes post infection to detect PLE circularization. Strains complemented with a plasmid harboring *int* under control of an IPTG inducible P_{tac} promoter or the empty vector (EV) control were induced 20 minutes prior to phage infection. **e**, Western blot for PLE 1 harboring endogenously FLAG-tagged Int shows Int expression independent of ICP1 infection. Samples of PLE 1 FLAG-Int or untagged PLE 1 Int were collected at the indicated timepoints after infection by ICP1. **f**, PLE circularization is not necessary to block ICP1 plaque formation. Tenfold dilutions of ICP1 were spotted onto the listed *V. cholerae* lawns to observe the ability to form plaques (dark spots, zones of killing). ICP1 is unable to form plaques on PLE 1 even in the absence of *int*, while the CRISPR proficient phage is able to overcome PLE 1 and form plaques[24].

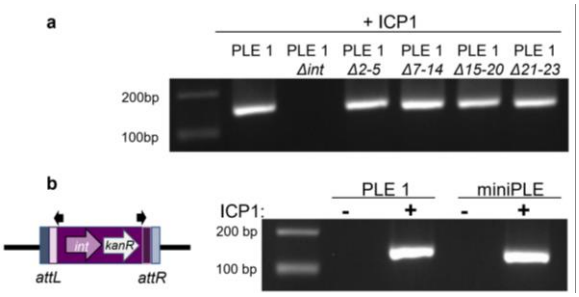


Figure 1.2. PLE 1 does not encode an RDF. a, Circularization PCR of PLE 1 variants lacking ORFs as indicated, taken 5 minutes after infection by ICP1 at an MOI of 5. **b,** Cartoon of the miniPLE (left) containing PLE 1 *int* and a kanamycin cassette (*kanR*) integrated into the *V. cholerae* chromosome in the same location as PLE 1 with the same *att* sites. Circularization of miniPLE (right) can be detected 5 minutes post ICP1 infection using the same primers as used for PLE 1 circularization.

(*attL*) and right (*attR*) attachment sites, resulting in excision of DNA between these sites (Fig. 1.1b). In addition to what has been documented regarding PLE circularization following ICP1 infection under laboratory conditions[22,24], we note that ICP1-dependent PLE circularization can be detected in cholera patient stool (Fig. 1.1c), underscoring that *V. cholerae* PLE responds to ICP1 infection during disease in humans.

To determine if Int plays a role in PLE circularization during ICP1 infection, PLE 1 *Δint* was challenged with ICP1. Unlike wild-type PLE, circularization was not detected in the PLE 1 *Δint* strain. PLE circularization was restored with *in trans* Int expression, but only during ICP1 infection (Fig. 1.1d, Supplementary Fig. 1.1a). As Int is necessary for PLE circularization to occur, we next investigated the expression pattern of Int during ICP1 infection. We introduced a FLAG-tag into the endogenous copy of *int* and confirmed that FLAG-tagged Int retained the ability to catalyze circularization within five minutes of ICP1 infection (Supplementary Fig. 1.1b). We detected FLAG-Int in uninfected cells and observed that the level of Int did not increase during ICP1 infection (Fig. 1.1e), showing that ICP1 infection does not induce *int* expression. Interestingly, although PLE 1 *Δint* cannot functionally circularize following phage infection, PLE 1 *Δint* still inhibits ICP1 plaque formation (Fig. 1.1f), suggesting that excision is induced separately from other components of PLE that are needed for anti-phage activity.

To determine if PLE 1 Int is a functional LSR, we performed integration assays to probe the ability of Int to recombine chromosomal *attC* and PLE *attP* sites. Through both *in vivo* (Supplementary Fig. 1.1c) and *in vitro* (Supplementary Fig. 1.1d) assays, we found that Int was sufficient to recombine *attP* and *attC* sites, as is characteristic of LSRs[39,40]. These assays demonstrate that Int is necessary for PLE circularization in response to ICP1 infection and that Int is a functional LSR that can catalyze recombination between *att* sites.

PLE requires another factor to direct circularization

As Int is constitutively expressed and not sufficient to catalyze the circularization of PLE 1 in the absence of phage infection (Fig. 1.1e, Supplementary Fig. 1.1a), we hypothesized that an RDF was required to direct Int to recombine the *attL* and *attR* sites as is characteristic of LSRs[42,43]. There are no conserved sequence characteristics of RDFs that enable homology-based identification[44], however, in characterized LSR/RDF systems of temperate phages, both the LSR and RDF are encoded within the same genome[42,45]. To evaluate if the RDF is PLE-encoded, PLE 1 strains harboring gene cluster deletions of predicted open reading frames (ORFs) were screened for circularization defects during ICP1 infection. Unexpectedly, all of the PLE ORF knockouts still circularized, implying that the RDF is not PLE-encoded (Fig. 1.2a). To

establish the minimal PLE-encoded factors required for circularization, we constructed a ‘miniPLE’, which has Int under control of its endogenous promoter and a kanamycin cassette, flanked by *att* sites, and integrated into the *V. cholerae* chromosome in the same location as PLE 1 (Fig. 1.2b). In support of the mutational analyses (Fig. 1.2a), the miniPLE circularized and excised from the chromosome during ICP1 infection (Fig 2b., Supplementary Fig. 1.2). Together with the inability of PLE 1 *Δint* to circularize (Fig. 1.1d), these results demonstrate that Int is the only PLE gene that is necessary for PLE 1 circularization during ICP1 infection.

ICP1-encoded PexA is involved in PLE circularization

Due to the specificity of circularization during ICP1 infection, we hypothesized that ICP1 encodes a gene product that directs Int-mediated PLE circularization during infection. To identify this phage-encoded gene product, we screened for ICP1 mutations that abolished miniPLE circularization during infection. Through this screen, we identified a mutant phage that failed to circularize the miniPLE (Supplementary Fig. 1.3a). Sequencing revealed that this mutant phage had a deletion of *orf50* and *orf51* caused by recombination between *orf49* and *orf52* (sites 18761-19800 in ICP1_2004_A, Sequence ID HQ641354 [https://www.ncbi.nlm.nih.gov/nuccore/HQ641354]), leading to a unique fusion gene with a novel stop codon. We evaluated the phage-encoded gene products within this region to determine if one or more was responsible for directing PLE excision during ICP1 infection. Ectopic expression of each gene product revealed that only the hypothetical ICP1 gene product annotated as *orf51* (YP_004250992 [https://www.ncbi.nlm.nih.gov/gene/?term=YP_004250992]) was sufficient to induce PLE circularization in the absence of phage infection (Fig. 1.3a, Supplementary Fig. 1.3b). We constructed a clean deletion of *orf51* and found that it was indeed necessary for miniPLE circularization (Fig. 1.3b, Supplementary Fig 3c), though interestingly, *orf51* is dispensable for ICP1 plaque formation (Supplementary Fig. 1.3d). As this gene product is both necessary and sufficient for Int-mediated PLE excision, we named it phage-encoded excisionase (PexA). PexA is a small protein unique to ICP1 isolates that has no sequence similarity to known proteins. Consistent with the rapid kinetics of PLE 1 circularization[22], we found that PexA is expressed *de novo* within five minutes of ICP1 infection (Fig. 1.3c), leading us to hypothesize that PexA is hijacked by PLE 1 to function as the RDF for Int-mediated PLE excision.

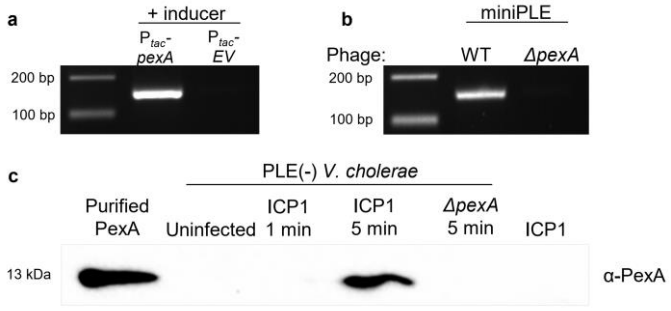


Figure 1.3. ICP1-encoded PexA is necessary and sufficient for PLE circularization. a, miniPLE circularizes in the absence of ICP1 infection upon ectopic expression of PexA. **b**, miniPLE does not circularize during ICP1 infection when *pexA* is knocked out. **c**, Western blot for PexA during ICP1 infection. PLE$^-$ *V. cholerae* cultures were probed for PexA at the listed times after infection with ICP1 or ICP1 *ΔpexA*. To determine if PexA is packaged in the phage particle, 5 times the PFU of phage as was used for infection was probed for the presence of PexA. Purified PexA (20 ng) was used as a positive control.

PexA is an RDF that directs Int to recombine *in vitro*

In order to direct recombination, characterized RDFs physically interact with their cognate LSR. The ability of PexA to physically interact with Int was probed with a bacterial adenylate

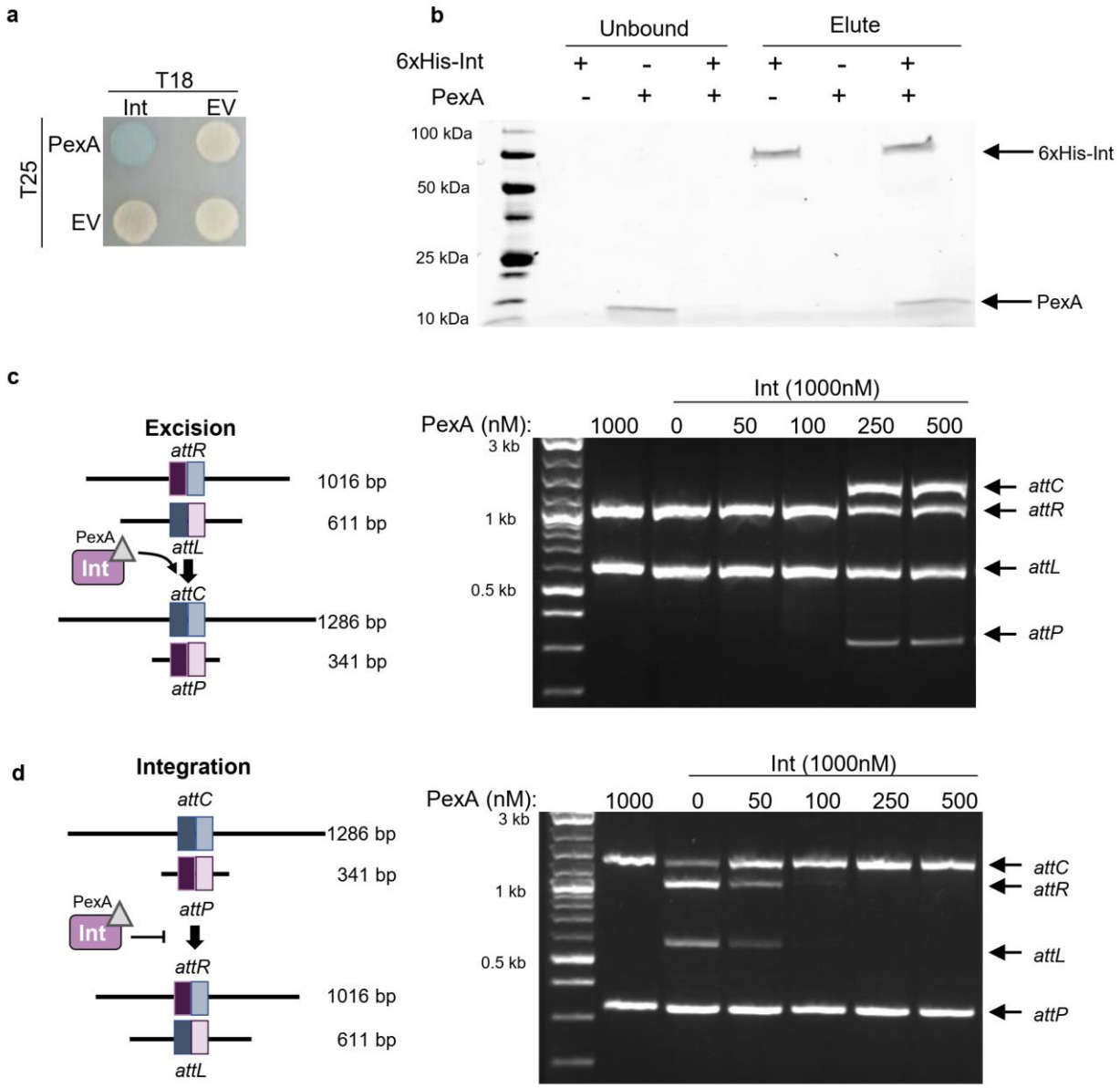


Figure 1.4. PLE exploits phage encoded PexA as the RDF for excision. a, Bacterial two hybrid analysis to detect protein-protein interactions. Cells containing fusions of PexA to the T25 subunit of adenylate cyclase *cyaA* (indicated on left) and Int to the T18 subunit (indicated on top) were spotted on X-gal. Blue colonies indicate a physical interaction between the proteins fused to the CyaA subunits. **b**, Purified 6xHisSUMO-Int (1mM) and/or PexA (0.25mM) were incubated for 30 minutes and then incubated with nickel resin. Unbound and eluted fractions were collected and were run on an SDS-PAGE gel. **c**, Left, cartoon of *in vitro* excision reaction shows dsDNA fragments containing *attL* and *attR* recombining to *attC* and *attP* in the presence of Int and PexA. Right, *in vitro* excision assay shows when purified PexA or Int are incubated alone with *attL* and *attR* (lanes 1 and 2, respectively), no recombination is detected. As the concentration of PexA increases, recombination products *attC* and *attP* are detected. **d**, Left, cartoon of *in vitro* integration reaction depicting the ability of PexA to inhibit the activity of Int when recombining *attC* and *attP* sites in an integration reaction. Right, *in vitro* inhibition of integration (right) by PexA can be seen by the loss of the *attL* and *attR* recombination products as purified PexA increases in concentration.

cyclase two-hybrid (BACTH) assay, in which LacZ expression was detected when both Int and PexA were fused to adenylate cyclase subunits (Fig. 1.4a). This interaction was further validated using an *in vitro* pulldown assay, in which PexA coeluted with 6xHis-tagged Int (Fig. 1.4b), showing that PexA can bind to Int *in vivo* and *in vitro*.

Recombination in characterized LSR/RDF systems requires solely the LSR, RDF, and DNA substrates[41,42,45]. To determine if PexA directs Int to excise and circularize PLE, *in vitro* excision assays were performed using PCR fragments containing the *attR* and *attL* sites and purified PexA and Int (Fig. 1.4c). Addition of neither PexA nor Int alone led to recombination between *attR* and *attL*; however, when Int and PexA were both added, recombination products were detected (Fig. 1.4c). RDFs have also been shown to block LSR-mediated integration[42]. Consistent with this model, we observed that addition of PexA to an *attC* and *attP* integration reaction blocks Int-mediated recombination *in vitro* (Fig. 1.4d). These data demonstrate that phage-encoded PexA is the RDF for PLE 1 Int and provide the first example of an LSR/RDF pair being encoded in different genomes.

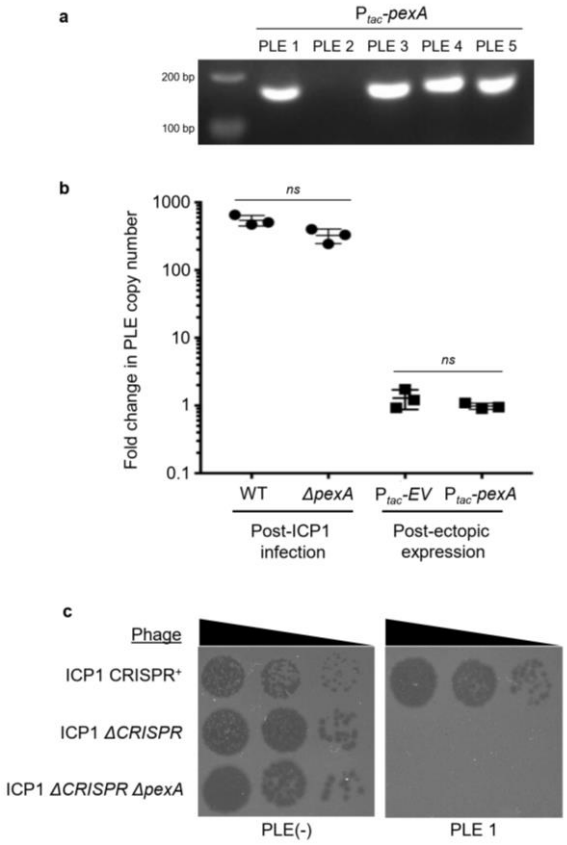


Figure 1.5. PexA directs historic PLEs to circularize and is dispensable for PLE 1 anti-phage activity. a, Circularization PCR of all PLEs probed during ectopic expression of PexA. **b,** Fold change in PLE 1 copy number as measured by qPCR. Left, PLE copy number was compared before and 20 minutes after ICP1$_{2004}$ or ICP1$_{2004}$ *ΔpexA* infection. Right, PLE copy number was examined before and 15 minutes following ectopic expression of an empty vector (EV) *or pexA*. Bars represent the mean and standard deviation of three independent replicates tested (*ns,* not significant by T Test). **c**, Tenfold dilutions of ICP1 spotted on *V. cholerae* PLE$^-$ or PLE 1 lawns showing the ability of different phage strains to form plaques (dark spots, zones of killing). ICP1 and ICP1 *ΔpexA* are unable to form plaques on PLE 1, while the CRISPR proficient phage is able to overcome PLE 1 and form plaques[24].

PexA is conserved and directs excision of historic PLEs

Analysis of ICP1 genomes from a 12-year period shows that PexA is maintained and that it is 99% identical in all ICP1 isolates (Supplementary Fig. 1.4a). The conservation of PexA leads us to speculate that although PexA is not essential for plaque formation under lab conditions (Supplementary Fig. 1.3d), it is likely integral to the ICP1 lifecycle in nature, and hence it may be hijacked as the RDF to signal ICP1 infection in all PLEs. Further, all five PLEs contain a putative LSR and respond to ICP1 infection by circularizing following infection[22], therefore we examined the conservation of the Int/PexA interaction. Upon ectopic expression of PexA, all PLEs, except PLE 2, demonstrated functional circularization (Fig. 1.5a). PLE 2 has the most diverse Int, sharing 25.7% amino acid identity with PLE 1 Int across 63% of the protein, while Int from PLE 3, PLE 4, and PLE 5 are more similar to

PLE 1 Int (Supplementary Fig. 1.4b). Consistent with the divergence of PLE 2 Int, PLE 2 also integrates into a unique site in the *V. cholerae* small chromosome[22] (Supplementary Fig. 1.4c), indicating that PLE 2 Int recognizes different *att* sites from the other PLEs. Altogether, this data indicates that PLE 2 evolved to recognize a unique RDF, possibly altering *att* site specificity in the process.

Although PexA appears to be stably maintained in the natural ICP1 isolates in our collection, the observation that PLE 2 Int has evolved to use a unique RDF indicates that PexA may not have always been a reliable cue enabling PLE to respond to ICP1 infection throughout their co-evolutionary history. We also found that PLE blocks ICP1 infection independent of Int and subsequent circularization (Fig. 1.1f), suggesting that PLEs may have evolved to use multiple phage products to induce activity. This model of PLE induction is in stark contrast to characterized PICI systems in which a single phage product is sufficient to activate the entire PICI excision-replication-packaging program[6,7]. To test if PexA plays an additional role in PLE activation, we analyzed PLE copy number for PexA-linked effects. During infection with ICP1 *ΔpexA*, PLE still replicated to high copy, although we did observe a mild but not significant defect relative to infection with wild-type phage (Fig. 1.5b), presumably because PLE is unable to excise from the chromosome and replicate quite as efficiently. Additionally, ectopic expression of *pexA* was not sufficient to drive PLE replication in the absence of ICP1 infection (Fig. 1.5b), thus PexA functions as the RDF for PLE-encoded Int to stimulate PLE excision but PexA does not appear to play a role in inducing other aspects of PLE activity. Consistent with PexA serving solely as the RDF and our previous observation that PLE *Δint* still blocks ICP1 plaque formation (Fig. 1.1f), ICP1 *ΔpexA* is still blocked by PLE 1 (Fig. 1.5c), confirming that PLE circularization is not required for PLE-mediated anti-phage activity.

Discussion

We report the first LSR/RDF system in which the interacting LSR and RDF are encoded in separate genomes. The exploitation of PexA by PLE demonstrates co-evolution with ICP1, as PLE has evolved to use a conserved phage protein to direct PLE excision and circularization. By constitutively producing Int, *V. cholerae* PLE 1 is perpetually ready to bind PexA, which is expressed early in infection, allowing for rapid response to ICP1 infection. As PexA is unique to ICP1, its function as the RDF for PLE 1 Int defines one component underlying the molecular specificity of the interaction between PLE and ICP1. This interaction has largely persisted, despite the spatiotemporal distribution of samples from which PLEs have been isolated. For example, the oldest *V. cholerae* isolate known to harbor PLE 5 is an Egyptian isolate from 1949[22], yet PLE 5 is specifically mobilized by PexA (Fig. 1.5a). Conversely, PLE2 has evolved to recognize a unique RDF while maintaining its response to ICP1. This switch may have been possible because PexA is not essential for ICP1, and although not represented in our collection, it may have been lost in historical epidemics, forcing PLE to co-evolve to continue to recognize persistent ICP1 infection. This specificity and adaptability supports the hypothesis that PLE is not a general phage defense mechanism, but an evolved, highly attuned, and specific response evolved to combat continued predation by ICP1.

Since the capacity of PLE to block ICP1 plaque formation is not dependent on PexA-mediated PLE circularization during infection (Fig. 1.5c), we hypothesize that other, perhaps several other, ICP1-encoded signals induce PLE activity. This model is divergent from the characterized PICIs in Gram-positive bacteria, in which a single phage protein is necessary to de-represses a PICI master repressor. While PICIs can evolve to recognize a variety of helper phage inducing proteins[8], only a single input is required, leading to induction of the PICI-encoded RDF, and ultimately the entire excision, replication, and packaging cycle[5–7]. Accordingly, phages that avoid inducing PICI activity can be selected for *in vitro*[46]. Not only is the constitutive expression of PLE 1 Int in contrast to the general regulation pattern seen in PICIs and temperate phage alike, our findings also indicate that single evolutionary events that compromise PLE inducing genes in ICP1 cannot prevent PLE induction entirely, thus revealing the evolutionary pressures that lead to the apparent fixation of an active anti-PLE CRISPR-Cas system in ICP1 isolates[24,25].

PLE excision during ICP1 infection enables mobilization of PLE to neighboring cells, and we have previously shown that PLE transduction is Int dependent[22]. It is interesting to note, then, that PLE exploitation of PexA forces ICP1 to directly contribute to horizontal spread of this anti-ICP1 genomic island. Beyond lysogenic phage, serine recombinases have also been shown to have roles in mobilization of pathogenicity islands and antibiotic resistance cassettes, such as the tandem LSRs that control integration and excision of SCC*mec*, which confers methicillin resistance to *S. aureus*[47]. Since many genes encoded in PLEs have unknown functions and PLE$^+$ *V. cholerae* isolates have been observed with increasing prevalence during recent epidemic sampling[25], it is possible that PLEs may have an underappreciated role in cholera pathogenesis and disease. PLEs confer a known fitness advantage in defending against ICP1 predation[22,24], however in light of some of the functional similarities between PLEs and SaPIs[22,34] it is important to consider that PLEs may confer additional traits relevant to the molecular epidemiology of contemporary clones. As such, it is important to understand the mobilization of such islands, which we note, cannot be achieved through genomic analyses alone as there is no sequence-based signature that implicates a role for ICP1 in PLE mobilization. Uncovering the RDF activity of PexA provides insight into the multitude of mechanisms by which newly discovered and even well characterized anti-phage islands can be mobilized by their target phages. Like many lytic phages, ICP1 is being considered for its therapeutic utility[48]; however, without mechanistic insight into phage-bacterial interactions in nature, we may overlook the important ways in which ICP1, as well as phage in general, contribute to pathogen evolution.

Methods

General growth conditions.

Strains, primers, plasmids, and phage used in this study are listed in Supplementary Tables 1.1-1.5. Strains were grown with aeration in LB (lysogeny broth) or on LB agar plates at 37°C, unless otherwise noted. Where necessary, cultures were supplemented with streptomycin (100 μg/mL), spectinomycin (100 μg/mL), kanamycin (*V. cholerae* 75 μg/mL, *E. coli* 50 μg/mL), ampicillin, (*V. cholerae* 50 μg/mL, *E. coli* 100 μg/mL), chloramphenicol (*V. cholerae* 2.5 μg/mL, *E. coli* 25 μg/mL), or 5-bromo-4-chloro-3-

indolyl-beta-D-galacto-pyranoside (X-gal), 40 μg/mL. Ectopic expression constructs in *V. cholerae* were induced with 1 mM IPTG and 1 mM theophylline (P_{tac}) or 0.1% arabinose and 1 mM theophylline (P_{bad}). Phage were propagated on *V. cholerae* hosts using the soft agar overlay method[22]. High titer phage stocks were collected by polyethylene glycerol precipitation[49].

Strain construction.
PCR products to make chromosomal *V. cholerae* mutants, including chromosomal expression constructs, were created by SOE (splicing by overlap extension) PCR and introduced by natural transformation[50]. Primer sequences are shown in Supplementary Table 1.2. Ectopic P_{tac} gene expression vectors were created from a modified pMMB67EH vector[51] or pBAD vector engineered to contain a theophylline inducible riboswitch (Riboswitch E)[52]. Plasmids were constructed using Gibson assembly (NEB), and introduced into *V. cholerae* through conjugation with *E. coli* S17. ICP1 mutants were created as described through CRISPR-Cas gene editing[53]. Briefly, an IPTG-inducible type 1E CRISPR-Cas system was engineered in *V. cholerae* and used to target various regions of the ICP1 genome. Escape phage that are able to form plaques on the targeting host have acquired random mutations in the protospacer region, often as a result of recombination events leading to random deletions. To generate clean deletions, a targeting plasmid was engineered with an additional editing template that served as a recombination template to delete only the gene product targeted by the *V. cholerae* spacer. Successful recombination events allowed ICP1 to form plaques on the targeting host, and all clean deletion constructs were confirmed with DNA sequencing over the region of interest.

Circularization and excision PCRs.
Stool specimens were collected and stored from previous studies[24]. Total DNA was extracted from 100 μl stool samples using the DNeasy blood and tissue kit (Qiagen), and 2 μL of extracted DNA was used as template to detect PLE, ICP1, and circularized PLE by PCR. For detection of PLE circularization during phage infection, *V. cholerae* strains were grown to OD_{600}=0.3, infected with phage at a multiplicity of infection (MOI) of 5, and allowed to incubate at 37°C with aeration. Samples were taken 5 minutes post infection, boiled for 10 minutes, and 2 μL was used as template for PCR using primers depicted in Fig 1b. Resulting reactions were run on a 2% agarose gel and imaged with Gel Green. For Int complementation, ectopic Int was induced for 20 minutes prior to phage infection. To detect PLE excision, 6 ng gDNA from uninfected *V. cholerae* and infected *V. cholerae* harvested 15 minutes post infection were used as templates for PCR with primers located in the *V. cholerae* chromosome flanking PLE 1. To detect PLE circularization following induction of ectopically expressed PexA, PLE^+ derivatives of *V. cholerae* at an OD_{600}=0.3 were induced for 5 minutes and samples were boiled and processed for PCR as described above. Images are representative of at least two independent experiments.

In vivo recombination assay.
Constitutively expressed *lacZ* from *E. coli* was engineered such that *lacZ* was flanked by 300 bp containing att_P from circularized PLE and 70 bp containing att_C from a *V.*

cholerae repeat (VCR). This construct was integrated into the *V. cholerae* genome at a fixed position (VC2338) by natural transformation. A plasmid with Int or the empty vector control was mated into the reporter strain and individual colonies were picked into 1 mL LB and 2 μL was spotted onto indicator plates with X-gal, IPTG and theophylline.

PLE circularization screen with mutant phage.
Mutant phages, created by targeting the ICP1 genome with CRISPR and collecting viable escape phage[53], were used to infect the miniPLE *V. cholerae* host using the soft-agar overlay method. Plaques were picked into 50 μL water and boiled for 10 minutes, and 2 μL of the boiled template was used for circularization PCR as described above. Mutant phage that did not circularize miniPLE were sequenced, and, since mutants were not clean deletions of individual gene products, the individual gene products that were missing or mutated were cloned into an inducible vector to test for miniPLE circularization. Strains grown to OD_{600}=0.3, induced for 20 minutes, and boiled and processed for PCR as described above. Images are representative of at least two independent experiments.

Western Blot.
V. cholerae cultures were grown to OD_{600} = 0.3 and infected with the indicated phage at MOI = 5, and 2 mL samples were taken at the time points indicated. Samples were washed with methanol and phosphate buffered saline (PBS) before being pelleted, re-suspended in 1x Leammli buffer (Bio-rad), and boiled for 10 minutes. Boiled samples were then loaded onto an SDS-PAGE gel for blotting analysis. Rabbit-α-FLAG primary antibodies (Sigma) were used at a dilution of 1:3000 to detect endogenous FLAG-Int, and custom rabbit-α-PexA primary antibodies (GenScript) were used at a dilution of 1:15000 to detect PexA. Both were incubated with goat-α-rabbit-HRP conjugated secondary antibodies at a dilution of 1:5000 (Bio-rad). Blots were developed with Clarity Western ECL Substrate (Bio-rad) and imaged on a Chemidoc XRS Imaging System (Bio-rad). Images are representative of at least two independent experiments.

Quantitative PCR.
To examine PLE replication during phage infection, *V. cholerae* samples were grown to an OD_{600}=0.3, split into two tubes, and incubated with the listed phage at MOI= 2.5 for 20 minutes at 37ºC with aeration. Samples were taken just before phage infection and 20 minutes following infection, boiled for 10 minutes, and diluted 1:1000. The qPCR reaction was run using primers Zac14 and Zac15 (Supplementary Table 1.1) and iQ SYBR Green Supermix (Bio-Rad) and were run on a CFX Connect Real-Time PCR Detection system (Bio-Rad) with the following conditions: denaturation at 95ºC for 5 minutes, followed by 40 cycles of 95ºC for 10 seconds and 60ºC for 30 seconds. To verify the results, a melt curve analysis was additionally performed. Results were analyzed with 2-tailed paired t test. To examine the effect of ectopic *pexA* expression on PLE replication, *V. cholerae* strains were grown to an OD_{600}=0.3 and induced for 15 minutes at 37ºC with aeration. Samples were then taken and boiled for 10 minutes and compared to samples taken immediately before induction and run as described above. Results were analyzed with a 2 tailed unpaired t test. All experiments were done with

three independent biological replicates and each template was quantified in technical duplicate.

Bacterial adenylate two-hybrid (BACTH) assay.
Genes of interest were cloned into the multiple cloning sites of the pUT18 and pKT25[54] using Gibson Assembly. Three independent colonies were separately picked into 1 mL of LB and 3 μL was spotted onto selective medium containing kanamycin and ampicillin with 0.5 mM IPTG and X-gal. Plates were incubated for 24 hours at 30°C before being imaged.

Protein preparation.
E. coli BL21 cells containing a pE-SUMO fusion to the construct of interest (cloned using Gibson Assembly), were grown in LB with antibiotics at 37°C to OD_{600}~0.9, and induced for 3 hours with 0.5 mM IPTG. Cells were centrifuged and resuspended in lysis buffer (50 mM HEPES pH 7.2, 300 mM NaCl, 20 mM imidazole, Pierce™ Protease Inhibitor Mini Tablets (Thermo), 1 mM TCEP, 0.5%Tx-100) and sonicated. Cell debris was removed by centrifugation (18,000x g for 40 minutes at 4°C), and the lysate was applied to a Nickel resin affinity column (HisPur Ni-NTA Resin). The column was washed with two column-volumes wash buffer (50m M HEPES pH 7.2, 1 M NaCl, 20 mM Imidizole, 1mM TCEP) and eluted with elution buffer (50m M HEPES pH 7.2, 300 mM NaCl, 300 mM Imidizole, 1 mM TCEP). Eluted 6xHisSumo-Int was then run through a HiTrap Heparin HP 5 mL column, and pooled fractions were run on a Superose 6 Increase 10/300 GL column on an AKTA Pure 25L system (GE Healthcare). Eluted 6xHisSumo-PexA was run on a Superose 6 Increase 10/300 GL column. To cleave the SUMO tag, 1 μL SUMO protease was added per 100 μg of protein and incubated overnight at 4°C. The mixture was then bound to Novex His-Tag Dynabeads and the unbound fraction was collected and analyzed by SDS-PAGE visualized with *Stain-Free* technology (Bio-rad).

In vitro pulldown.
Purified 6xHis-SUMO-Int (1000 nM) and/or untagged PexA (250 nM) were added to Novex Dynabeads His-Tag Isolation & Pulldown beads with Binding/Wash buffer (50 mM Sodium-phosphate pH 8, 300 mM NaCl, 0,01% Tween-20) and incubated rocking at room temperature for 10 minutes. Unbound protein was collected and the resin was washed 4x with Binding/Wash buffer. His Elution buffer (300 mM imidazole, 50mM sodium-phosphate pH 8, 300 mM NaCl, 0.01% Tween-20) was added, incubated for 5 minutes rocking at room temperature, and collected. Fractions were analyzed by SDS-PAGE visualized with *Stain-Free* technology (Bio-rad). Images are representative of at least two independent experiments.

In vitro recombination.
Purified Int and/or PexA were added to 20 μL reactions in the concentrations listed (Fig. 1.4c, Fig 4d, Supplementary Fig. 1.1d) with 200 ng purified PCR products containing the indicated *att* sites in buffer (50 mM HEPES pH 7.2, 150 mM NaCL, 10% glycerol, 0.5 mg/mL BSA, 5mM spermidine). Reactions were incubated for 2 hours at 37°C followed by 10 minutes at 75°C to heat inactivate. The entire 20 μL reaction was then run on a

2% agarose gel and visualized with Gel Green. Images are representative of at least two independent experiments.

Phage plaque spot plates.
V. cholerae grown to mid-log was added to 0.7% molten LB agar, poured over a solid agar plate, and allowed to solidify. 3 μL of each ten-fold dilution of phage in LB was spotted onto the solid surface and allowed to dry. Plates were incubated at 37°C overnight before being imaged. Images are representative of at least two independent experiments.

Genomic analysis.
Structure prediction of PLE 1 Int was performed using NCBI Conserved Domain Database[55] and the MPI bioinformatics toolkit[56]. PexA sequence was displayed with Weblogo[57]. Alignments for PLE Int were performed with PRALINE[58], and EMBOSS Needle was used to compare PLE 1 and PLE 2 Int[59].

Data availability. The data supporting the findings of the study are available in this article and its Supplementary Information files, or from the corresponding author upon request.

Acknowledgements: We would like to thank members of the Seed Lab for thoughtful discussions and feedback. Thanks to Dr. Lindsay Matthews and Dr. Lyle Simmons for the BACTH materials and bacterial strains and plasmids used for protein purification. Thanks to the Glaunsinger Lab for use of their AKTA and specifically Matt Gardner for training on the system. This work was supported by a Kathleen L. Miller Fellowship to ACM from the Henry Wheeler Center for Emerging and Neglected Diseases and grant R01AI127652 to KDS from the National Institute of Allergy and Infectious Diseases (US). KDS is a Chan Zuckerberg Biohub investigator.

Chapter 3

Phage defense island hijacks phage-encoded accessory helicase to drive replication

Amelia McKitterick, Stephanie Hays, and Kimberley D. Seed

Summary

In order to successfully infect, viruses have evolved a variety of different mechanisms to sabotage their host and redirect host machineries to facilitate viral propagation. Phages, viruses of bacteria, are masters of host takeover, with a well-characterized propensity to degrade the host cell's chromosome to recycle nucleotides into the newly synthesized phage genomes. Successful mechanisms to defend against phage are primarily found on mobile defense islands, which are predicted to help facilitate horizontal spread of defense related genes. One such island in *V. cholerae*, referred to as PLE, is deployed to specifically defend *V. cholerae* against the prevalent phage, ICP1. PLE specifically responds to infection by ICP1 by hijacking a phage protein to direct PLE excision from the *V. cholerae* chromosome. Following excision, PLE replicates and is packaged for transduction while concurrently inhibiting phage production. Here, we demonstrate the mobility of PLE out of the *V. cholerae* chromosome is necessary to mediate successful defense against ICP1 and escape phage-encoded nucleases that degrade the host chromosome. We show that PLE hijacks a phage-encoded accessory helicase to directly facilitate PLE replication, which additionally contributes to PLE-encoded gene dosage. We found that this helicase is not essential for ICP1 replication but does contribute to phage fitness. To evaluate the conservation of this helicase, we analyzed phage recovered from cholera patient stool samples and discovered ICP1 isolates with an alternative helicase allele at the same locus. The two helicase alleles were each identified in phage from two geographically distant sites within Bangladesh where cholera is endemic, suggesting that regional differences in PLE could drive variation in ICP1 genomic content. Despite differences in the two helicases, PLE is able to use either ICP1-encoded helicase to drive PLE replication, underscoring the importance of mobilization for defense islands to themselves escape virally encoded host takeover.

Introduction

Viruses and mobile genetic elements (MGEs) have been found to be associated with organisms from all branches of the tree of life [60]. In order to successfully infect their host, viruses elaborate a variety of host-takeover programs that inhibit host activities while promoting viral processes. Bacteriophages, or phages, are viruses that infect bacterial hosts and have profound effects on human health and disease, as well as bacterial fitness [2,61]. Of interest, lytic phages, which infect and kill their bacterial hosts within a single round of infection, have recently come to light as having impactful roles in mediating the makeup of bacterial populations, such as the microbial composition of the human gut microbiome [17], and as biocontrol agents for antibiotic resistant infections [62].

Lytic phages are particularly insidious to their bacterial hosts—upon infection, phages, like the *Escherichia coli* phage T4, can express a variety of genes that mediate host-cell takeover programs. T4 expresses genes that shut down and redirect host transcriptional machinery to favor transcription of phage genes, as well as nucleases that degrade the host chromosome to inhibit host gene expression as well as free up nucleosides that are incorporated into the rapidly replicating phage genome [63–65].

Paradoxically, phages also contribute to bacterial population diversity and complexity by facilitating horizontal gene transfer (HGT) [2,32]. In addition to generalized and specialized transduction, which are well characterized mechanisms by which phages can spread bacterial genetic material to neighboring cells [66], additionally lytic phages can facilitate the spread bacterial plasmid DNA from lysed cells to neighbors, increasing the dynamic range of genetic material that can be shared within a population [15].

In order to defend against phages, bacteria have evolved to encode a wide variety of anti-phage activities that have varied mechanisms of action [30,31]. Widely characterized restriction-modification systems and CRISPR-Cas systems inhibit phage through targeted cleavage of the infecting phage genome while toxin/antitoxin systems and abortive infection systems function through killing of the infected host cell [26,30,31]. Phage parasites, such as phage inducible chromosomal islands (PICIs), are another defense mechanism that protects the bacterial host while also exploiting the phage replication program to facilitate spread [5]. One type of PICI, the well characterized *Staphylococcus aureus* pathogenicity islands (SaPIs), are induced by infection with a helper phage, compete over host replication resources, and steal phage packaging proteins to selfishly package the SaPI genome, which negatively impacts the ability of the helper phage to complete its lifecycle [7,12,34]. Despite diverse mechanisms, phage defense programs must somehow overcome phage-mediated host takeover and go on to prevent rampant phage propagation through the bacterial community. Genomic analyses to localize anti-phage mechanisms in bacterial genomes have revealed that they tend to cluster together what is referred to as defense islands (DIs) [67]. Analysis of DIs has even led to the discovery of new phage defense mechanisms that have yet to be further elucidated solely due to the prevalent clustering of these defense systems on MGEs [68]. While hypothesized to have roles in HGT, the prevalence of phage defense systems on genomic islands has yet to be explained.

Vibrio cholerae, the etiological agent of the diarrheal disease cholera, is constantly under assault by phages both in aquatic environments as well as in human hosts [18,19,36]. The most dominant phage that preys on epidemic *V. cholerae* is ICP1, a lytic myovirus that is consistently isolated from cholera patient stool samples in regions where cholera is endemic, such as from Dhaka, Bangladesh [19,21,69]. In response to the consistent assault from ICP1, *V. cholerae* has acquired the phage-inducible chromosomal island-like element (PLE), a highly specific phage defense island that blocks plaque formation by ICP1 [22]. PLE excises from the host chromosome during ICP1 infection, replicates to high copy, and is specifically transduced to neighboring cells; concurrently, PLE replication negatively impacts ICP1 replication, contributing to the inhibition of ICP1 production observed [23]. PLE encodes a large serine recombinase, Int, that catalyzes the PLE excision and circularization reaction by physically interacting with ICP1-encoded PexA, a small protein of unknown function that is specific to ICP1 and is hijacked by PLE to act as a recombination directionality factor (RDF) [29]. Once excised, PLE begins to replicate and then is thought to steal structural proteins from ICP1 to facilitate its own transmission. Once packaged, PLE triggers accelerated lysis of the infected culture allowing for release of PLE transducing particles from the infected cells, ultimately killing the infected *V. cholerae* host but protecting the population as no infectious ICP1 progeny are produced [22]. Five PLEs have been identified in epidemic *V. cholerae* isolates. In addition to the ability in inhibit ICP1, PLEs are also characterized by conserved genomic architecture and the aforementioned PLE lifecycle during ICP1 infection [22].

Recent work has uncovered a PLE-encoded factor that is necessary for PLE replication: the replication initiation factor, *repA* [23]. Expression of *repA* is induced by ICP1 infection (unpublished), facilitating origin binding and recruitment of replisome proteins that have yet to be identified. In the absence of ICP1 infection, however, RepA is not sufficient to drive PLE replication [23]. PLE is not predicted to encode replication machinery, suggesting that other phage-encoded gene products, such as predicted ICP1 polymerase or helicase/primase, are required for PLE amplification. As all PLEs replicate following ICP1 infection [22], it stands to reason that the PLE has evolved to exploit conserved components of ICP1's replication machinery. Similar to PLE excision [22], PLE replication is essential for horizontal transmission of PLE transducing units, thus further underscoring the role of ICP1 in facilitating PLE HGT [23]; however, the relatively low rate of transduction suggests that robust PLE replication has other roles in PLE conflict with ICP1.

In order to inhibit ICP1, PLE must escape from ICP1-mediated host takeover during infection. While the mechanisms that ICP1 uses to overcome *V. cholerae* have not been characterized, ICP1 is able to quickly begin replicating its genome following infection [23] and produces about 100 virions within 20 minutes of infection [22]. Here, we identify ICP1 *ΔpexA* mutants that have the ability to escape PLE by acquiring mutations in the ICP1-encoded SF1B accessory helicase that we have named *helA*. We show that while this helicase is not necessary for ICP1 replication, it is essential for PLE replication during ICP1 infection. We go on to show that the excision- and replication-deficient PLE is susceptible to ICP1 host takeover through phage-encoded nucleases that degrade PLE while it remains integrated in the *V. cholerae* chromosome. Analysis of natural isolates of ICP1 from cholera patient stool samples revealed an alternative

SF1B helicase allele in phages shed from a rural hospital in Bangladesh, and functional comparison of the two alleles revealed that both alleles, though unrelated, can be hijacked by all PLEs to facilitate PLE replication. Though neither helicase is essential for ICP1, ICP1 faces impaired fitness in the absence of either accessory helicase, explaining their prevalence in ICP1 and *Vibrio* phages. PLE's capacity to use a variety of phage-encoded helicases to drive PLE replication underscores the critical role that replication plays in the PLE lifecycle, overall leading to a model in which inducible islands such as PLE excise from the host chromosome to escape from phage-mediated host takeover and to facilitate continued gene expression.

Results

ICP1 is able to escape excision deficient PLE by acquiring mutations in predicted helicase *helA*

Previous work has demonstrated the role for phage-encoded *pexA* in directing PLE excision during ICP1^{A} infection [29] (Fig. 2.1A). PLE mediated inhibition of ICP1 does not require PLE excision, so ICP1 *ΔpexA* is still blocked by PLE (Fig. 2.1B); however, ICP1^{A} *ΔpexA* is able to form rare plaques on PLE at a frequency of about 1 per 10^{6} phage (Fig. 2.1A, Fig. 2.1C). Due to the low efficiency of plaquing, we consider these phage to be "escape phage" that have acquired a mutation in the genome allowing them to overcome PLE that is deficient for mobilization. To identify the phage gene(s) that harbor mutations enabling escape, we collected and purified three escape phage and performed whole genome sequencing.

Analysis of the genomes of the escape phage revealed that all escape phage had acquired mutations in ICP1^{A} *gp147*, a predicted SF1B-type helicase which we have since named *helicase A* (*helA*) (Supplementary Table 2.1). SF1B type helicases are found broadly across all domains of life and include the well-studied *recD* and *pif1* [70]. In eukaryotes such as *Saccharomyces cerevisiae*, *pif1* has been implicated in telomere maintenance, Okazaki fragment processing, and resolution of G-quadruplex motifs [71], while *recD* is a core component of the *E. coli* RecBCD complex involved in DNA processing and repair [72]. Another prototypical SF1B type helicase is *dda,* encoded by phage T4, which is a non-essential accessory helicase that has been implicated in origin melting, translocating proteins off DNA, and a wide variety of other functions *in vitro*, although its exact role *in vivo* is unknown [73–75].

To validate the role of *helA* in ICP1^{A} escape from a mobilization-deficient PLE, we constructed a clean *helA* deletion and probed the mutant phage for the ability to overcome PLE. ICP1^{A}- encoded *helA* is not necessary for plaque formation on PLE (-) *V. cholerae*, and ICP1^{A} *ΔhelA* is still blocked by PLE, indicating that *helA* is not necessary for PLE induction (Fig. 2.1B). Similar to *pexA*, the absence of *ΔhelA* gives ICP1 an advantage on PLE (+) *V. cholerae*, allowing ICP1^{A} *ΔhelA* to form plaques at two orders of magnitude higher frequency than relative to ICP1^{A} *ΔpexA* on a PLE (+) host (Fig. 2.1C). Conversely, the double mutant ICP1^{A} *ΔpexA ΔhelA* is able to form small plaques on PLE at a relatively high efficiency (Figures 1B and 1C). ICP1^{A} *ΔpexA ΔhelA* plaques that form on PLE *V. cholerae* were picked and the plaquing efficiency was re-tested to determine if those phage were subsequently able to escape PLE at a higher rate (Supplementary Fig. 2.1). As these progeny phage re-plaqued at the same efficiency as ICP1^{A} *ΔpexA ΔhelA*, we conclude that they are not genetic escape phage

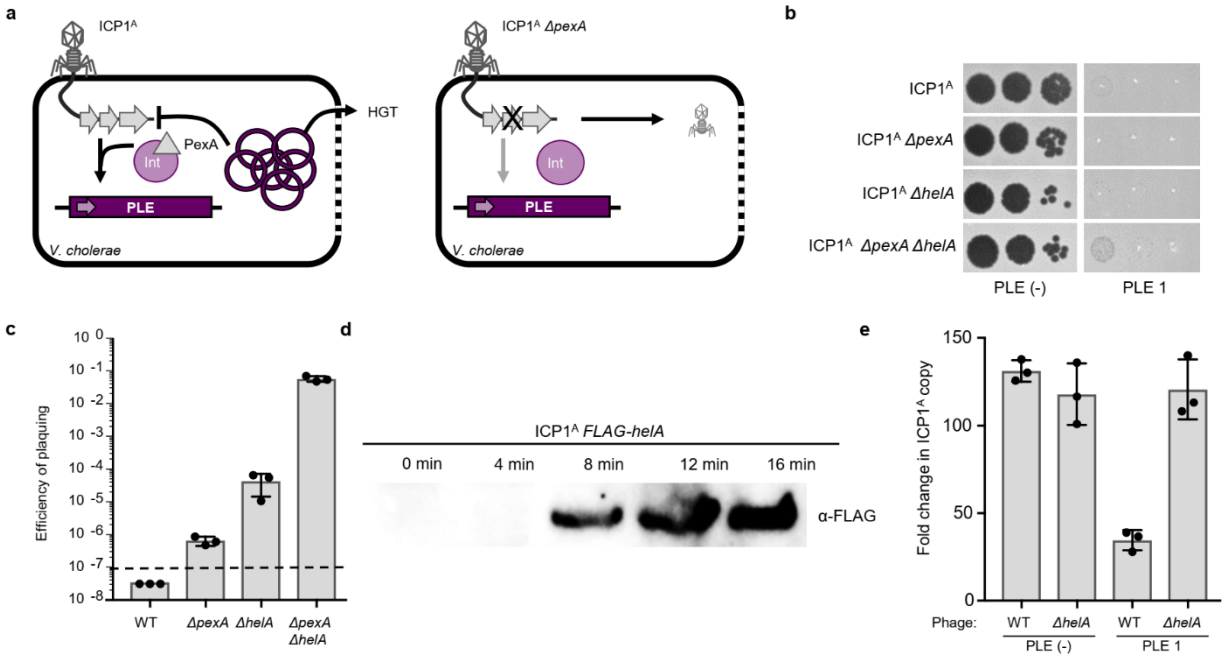


Figure 2.1. ICP1 is able to overcome excision-deficient PLE through loss of accessory helicase *helA*. a, Cartoon of the PLE response to ICP1 infection. Left, ICP1 infects PLE (+) *V. cholerae* and expresses PexA, which physically interacts with PLE-encoded Int to direct PLE circularization and excision. Excised PLE replicates to high copy number, inhibits ICP1 replication, and horizontally transduces to neighboring cells when the *V. cholerae* undergoes PLE-mediated accelerated lysis. Right, when ICP1 *ΔpexA* infects PLE (+) *V. cholerae*, PLE remains integrated in the host chromosome, and mutant phage are able to escape and form a plaque. **b,** Tenfold dilutions of ICP1 spotted on a *V. cholerae* lawn (grey). Zones of killing are shown in black. **c,** Efficiency of plaquing of ICP1 on a PLE (+) relative to a PLE (-) *V. cholerae* host. Dashed line indicates limit of detection. **d,** Western blot of endogenously FLAG-tagged HelA during infection of PLE (-) *V. cholerae*. **e,** Quantification of change in ICP1 genome copy number following 20 minutes of infection of the listed *V. cholerae* host as detected by qPCR.

but instead are able to overcome some aspects of a mobilization-deficient PLE through the loss of loss of ICP1^{A}-encoded *helA*.

We next wanted to characterize the role of *helA* for ICP1^{A} function. HelA is detectable in infected cells via Western blot within 8 minutes of ICP1^{A} infection (Fig. 2.1D), which is consistent with the onset of ICP1^{A} replication initiation [23], suggesting that *helA* may have a role in ICP1^{A} replication. As PLE has been demonstrated to diminish the level of ICP1^{A} replication [22,23], we hypothesized that PLE hijacks HelA during infection as a mechanism to interfere with the ability of ICP1^{A} to replicate. To test this hypothesis, we evaluated ICP1^{A} *ΔhelA* replication in the presence and absence of PLE by qPCR. In contrast to plaque formation, which requires multiple rounds of phage infection and replication to visualize a zone of killing, qPCR allows for measurement of phage replication in a single round of infection. Consistent with the ability to form a plaque, there are no deficiencies in ICP1^{A} *ΔhelA* replication relative to a wildtype phage over the course of the 20 minute infection cycle (Fig. 2.1E), indicating that *helA* is not essential for ICP1^{A} replication. Conversely, infection of a PLE (+) *V. cholerae* host with ICP1^{A} *ΔhelA* rescues ICP1 replication to the level that is observed in a PLE (-) host (Fig. 2.1E), suggesting that while not necessary for ICP1^{A}, *helA* is exploited by PLE to interfere with ICP1 during infection. However, because ICP1^{A} *ΔhelA* is not deficient for

replication in the absence PLE, the ICP1^{A} replication defect in the presence of PLE is not likely directly due to PLE-mediated hijacking of HelA activity.

ICP1-encoded *helA* is necessary for PLE replication

I CP1 and PLE replication appear to be inversely related, wherein ICP1 copy number is restored when PLE replication is abolished via deletion of the PLE origin of replication or RepA [23]. Therefore, the observed restoration in ICP1^{A} *ΔhelA* copy number during infection of a PLE (+) host implicated HelA in PLE replication. To test the role of *helA* in PLE replication, we infected PLE (+) *V. cholerae* with ICP1^{A} *ΔhelA* and monitored the change in PLE copy over the course of infection. While PLE is able to replicate to a high copy when infected with a wildtype phage, strikingly, PLE is unable to replicate in the absence of *helA* (Fig. 2.2A). This phenotype can be complemented by ectopic expression of *helA* during ICP1^{A} infection, supporting the role of *helA* in facilitating PLE replication.

SF1B-type helicases have been implicated in activities ranging from replication and genome maintenance to transcriptional regulation [71]. Additionally, the *S. aureus* phage parasites, SaPIs, make use of dUTPases as an anti-repressor to initiate the transcriptional program of the island, suggesting that these genomic islands can evolve to respond to phage-encoded proteins independent of their biological function for the phage [7,8]. As such, we next wanted to determine if *helA* had a direct role in PLE replication or if it was necessary to transcriptionally activate the island to allow for production of PLE-encoded proteins, such as *repA*, that are essential for PLE replication. To test the involvement of *helA* in PLE replication, we made use of a minimal PLE replication system referred to as the "midiPLE" [23]. The midiPLE contains only the endogenous PLE integrase as well as the PLE origin of replication, integrated in the same chromosomal location as PLE in the *V. cholerae* chromosome. This construct is competent to excise from the chromosome following *pexA* expression during ICP1^{A} infection, but is unable to replicate without ectopic expression of the PLE-encoded replication initiator, *repA*. When *repA* is provided *in trans*, midiPLE replicates

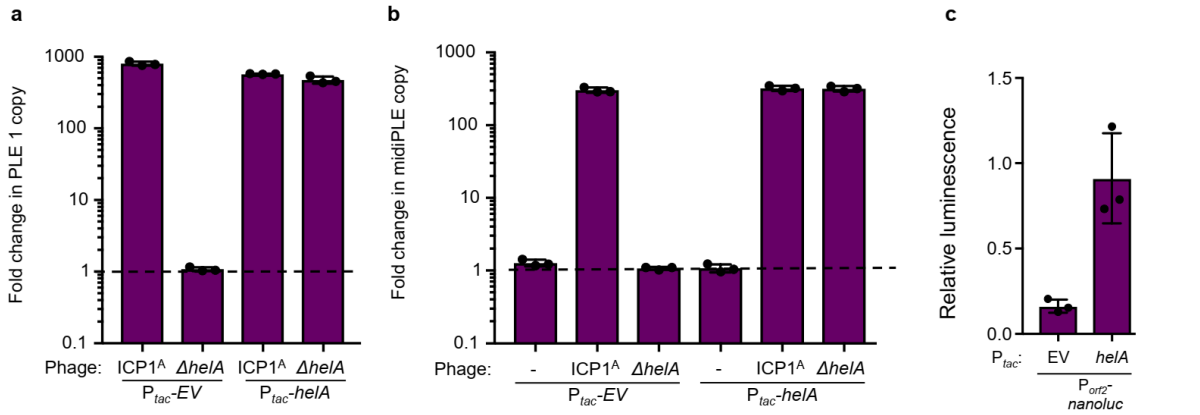


Figure 2.2. ICP1-encoded *helA* is necessary for PLE replication. a, Quantification of change in PLE copy number following infection by the listed ICP1 strain as measured by qPCR. Plasmids were induced 20 min prior to phage infection. Dashed line indicates no change in copy number. **b,** Quantification of change in miniPLE copy number following infection by the listed ICP1 as measured by qPCR. Ectopic *repA* and plasmids were induced 20 min prior to phage infection. **c,** Change in luminescence of P_{orf2}-*nanoluc* reporter following infection by ICP1^{A} *ΔhelA* relative to the change in luminescence following infection by ICP1^{A}.

during ICP1^{A} infection (Fig. 2.2B). In comparison to wildtype phage, midiPLE fails to replicate during infection with ICP1^{A} *ΔhelA*, and this phenotype can be complemented by expressing *helA in trans*, showing that *helA* is necessary for PLE replication independent of other PLE-encoded genes and supporting the conclusion that HelA is directly involved in PLE replication. Interestingly, *helA* is not sufficient to stimulate PLE replication in the absence of ICP1^{A} infection (Fig. 2.2B), indicating that other phage, or possibly *V. cholerae*, components are additionally required to facilitate PLE replication.

PLE replication contributes to anti-phage gene dosage

In the course of replication sampling during ICP1^{A} infection, we observed a defect in PLE-mediated accelerated lysis that correlated with a loss in PLE replication. A culture of PLE (+) *V. cholerae* infected with ICP1 typically lyses 20 minutes after infection, while an infected PLE (-) culture takes upwards of 90 minutes to lyse [22]. However, we observed that cultures infected with ICP1^{A} *ΔhelA* consistently had a delay in lysis, suggesting impaired PLE activity, and ectopic expression of *helA* led to intermediate lysis phenotypes (Supplementary Fig. 2.2A). Though the basis for PLE-mediated accelerated lysis is not known, we reasoned that robust PLE replication enhances expression of PLE-encoded genes merely through increasing the template copy number. To test this hypothesis, we created a nanoluciferase transcriptional reporter cloned downstream of PLE *orf2* (P_{orf2}*nanoluc*) to quantify defects in PLE transcription when PLE is unable to replicate (Supplementary Fig. 2.2B). Relative to infection with wildtype ICP1^{A}, P_{orf2}*nanoluc* produced 0.16 as much luminescence during infection with ICP1^{A} *ΔhelA* (Fig. 2.2C). When PLE replication was restored through ectopic expression of *helA*, the reporter activity resulting from infection with ICP1 *ΔhelA* was restored to wildtype levels, demonstrating that PLE copy number dictates the global level of PLE transcription. As such, inhibition of PLE replication leads to phenotypes such as delayed lysis during ICP1^{A} infection and potentially contributes to the ability of ICP1^{A} *ΔhelA* to escape PLE.

ICP1 overcomes replication and excision deficient PLE through phage-encoded nuclease

As ICP1-encoded *pexA* is necessary for PLE excision [29] and *helA* is necessary for PLE replication during ICP1 infection (Fig. 2.2A), we next wanted to understand how ICP1^{A} *ΔpexA ΔhelA* is able to overcome PLE (Fig. 2.1B). Even when PLE is challenged by ICP1^{A} *ΔhelA* and is unable to replicate, leading to transcriptional deficiencies, PLE is still able to excise from the *V. cholerae* chromosome and is more inhibitory than when it is maintained in the chromosome, leading us to speculate that the position of PLE in the cell, either intra- or extrachromosomal, is important for its activity. Phages are known to encode nucleases that attack the bacterial chromosome, freeing up nucleosides that can then be incorporated into the newly synthesized phage genome [64]. Additionally, deep sequencing of the total DNA in ICP1 infected *V. cholerae* cells shows that proportion of *V. cholerae* DNA decreases over the course of infection [23]. This observation led us to hypothesize that nucleolytic activity encoded by ICP1^{A}, deployed to degrade the *V. cholerae* chromosome during infection, is able to degrade PLE when PLE is stuck in the chromosome unable to replicate, allowing for ICP1^{A} to form some plaques on PLE (+) *V. cholerae*. To test this hypothesis, we made use of a minimal PLE

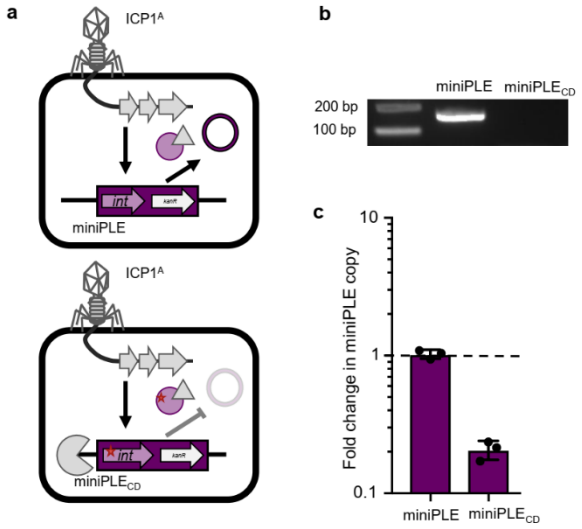


Figure 2.3. Excision and replication deficient PLE is susceptible to ICP1-encoded nucleases. **a,** Cartoon of miniPLE during ICP1 infection. Left, miniPLE-encoded Int (circle) is directed to excise miniPLE during ICP1 infection by ICP1-encoded PexA (triangle), leading to single-copy circularized miniPLE. Right, catalytically dead miniPLE$_{CD}$ Int (circle with red star) is unable to excise miniPLE during ICP1 infection, potentially rendering the miniPLE to phage-encoded DNAses (pac-man). **b,** Circularization PCR of listed miniPLE from boiled ICP1^{A} plaques. **c,** Change in miniPLE copy following ICP1^{A} infection as measured by qPCR.

system, the miniPLE, that has the PLE-encoded integrase but lacks an origin of replication (Fig. 2.3A). Thus during infection, the miniPLE is able to excise from the host chromosome and circularize, but is not able to replicate [29]. To simulate an excision-deficient miniPLE, we created miniPLE$_{CD}$, which possesses a point mutation in the catalytic serine residue in the miniPLE-encoded integrase, rendering the construct unable to excise from the chromosome (Supplementary Fig. 2.3). During the course of ICP1^{A} infection, the miniPLE successfully excises from the *V. cholerae* chromosome (Supplementary Fig. 2.3) and is maintained as a stable episome with no change in copy number (Fig. 2.3B). In comparison, the miniPLE$_{CD}$ that is unable to escape the *V. cholerae* host chromosome decreased in copy number during infection with ICP1^{A}, indicating that it is susceptible to ICP1^{A}-encoded nucleases. Thus, not only is PLE mobilization important for HGT [22,23], but it is also essential for PLE escape from ICP1 takeover of the *V. cholerae* host.

Diverse SF1B helicases are maintained in ICP1 and contribute to ICP1 fitness

Due to the importance of PLE replication in PLE gene dosage and avoiding ICP1-mediated host takeover, we next hypothesized that ICP1 would evolve to abolish PLE replication by accumulating mutations in the *helA* allele, indicative of co-evolution between the two entities. To identify signatures of co-evolution, we examined HelA from sequenced isolates of ICP1 that had been recovered from epidemic sampling in Dhaka, Bangladesh. HelA from ICP1 isolated from epidemic sampling from 2001 to 2017 is over 99% identical at the amino acid level indicating that there is either little pressure for HelA to evolve over time, or HelA mutations cannot be tolerated in nature (Supplementary Table 2.6). Even though there is no change in the ability of ICP1 to replicate in a single round of infection in the absence of *helA* (Fig. 2.1E), ICP1^{A} *ΔhelA* forms plaques that are on average 0.75 times smaller than wildtype phage plaques (Fig. 2.4A), showing that mutant phage are less fit in the absence of *helA* and supporting the notion that functional *helA* must be maintained by ICP1 in nature

Despite having a high degree of conservation, *helA* is not considered part of the core ICP1 genome [21]: two phage isolates recovered from cholera patient stool samples from 2006 in Dhaka do not encode *helA*, but instead have an alternative SF1B type

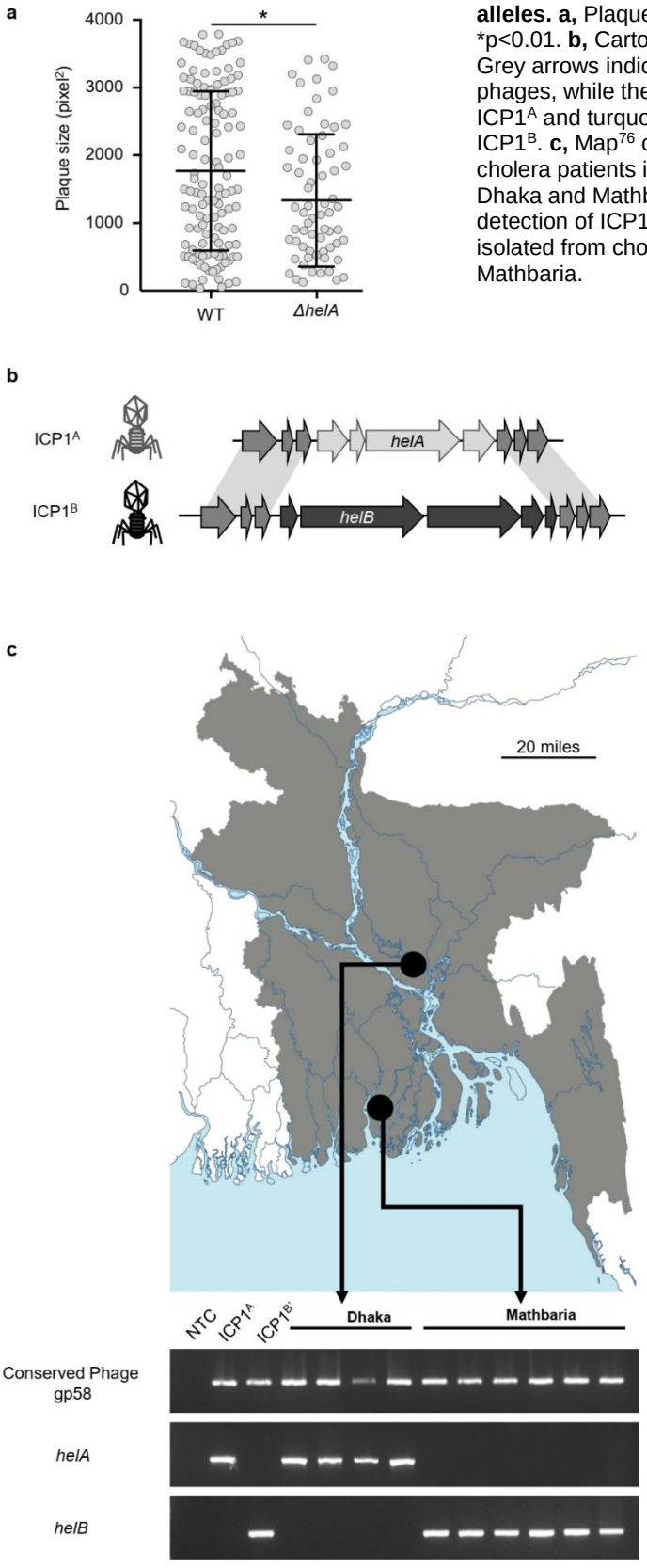


Figure 2.4. ICP1 encodes one of two accessory helicase alleles. a, Plaque size of listed phage on PLE (-) *V. cholerae*. *p<0.01. **b,** Cartoon of ICP1 accessory helicase locus to scale. Grey arrows indicate gene products shared between the two phages, while the mint arrows indicate gene products unique to ICP1^{A} and turquoise arrows indicate gene products unique to ICP1^{B}. **c,** Map[76] of distribution of ICP1 SF1B alleles shed by cholera patients in Bangladesh. Top, map of Bangladesh with Dhaka and Mathbaria marked. Bottom, agarose gel depicting PCR detection of ICP1 conserved gene *gp58*, *helA*, and *helB* in phages isolated from cholera patient stools collected in Dhaka or Mathbaria.

helicase in the same locus, which we call *helicase B* (*helB*) (Fig. 2.4B). HelB is 24% identical to HelA, with a conserved P-loop ATPase domain but HelB has an extended C-terminus that contains a domain of unknown function, DUF2493 (Supplementary Fig. 2.4A). In addition to having low sequence identity, *helA* and *helB* are flanked by different, unrelated genes each encoding products with no predicted structure or function (Fig. 2.4B), suggesting that ICP1 is unable to lose *helA* in an attempt to avoid PLE replication but instead that it must swap *helA* for another accessory helicase.

We then performed a BLASTP search of the National Center for Biotechnology Information's nonredundant protein sequence database to identify the origin of *helA* and *helB*. Homologs of HelA are commonly found in phages of marine bacteria, and, particularly, in a group of related myoviruses that are capable of infecting non-cholera *Vibrios* (Supplementary Fig. 2.4B). Of note, two of the *Vibrio* phages were also predicted to encode a homolog of one of the proteins flanking HelA in ICP1^{A} indicating that the *helA* allele could have been shared with a common ancestor of these phages. Conversely, HelB is more distantly related, with the only

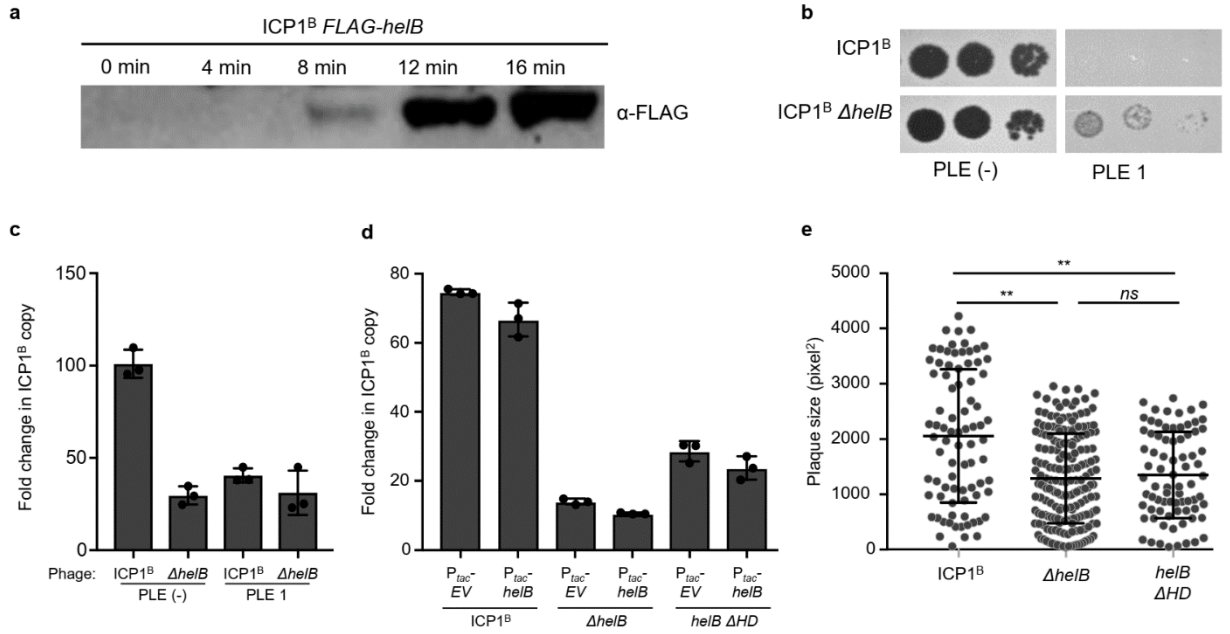


Figure 2.5. Loss of *helB* leads to a defect in ICP1 fitness. a, Western blot of endogenously FLAG-tagged *helB* at the listed time points following infection of PLE (-) *V. cholerae*. **b,** Tenfold dilutions of ICP1 spotted on the listed *V. cholerae* lawns. **c,** Fold change in ICP1 copy number following 20 minutes of infection of the listed *V. cholerae* host as measured by qPCR. **d,** Fold change in ICP1 copy number following 20 minutes of infection of the listed *V. cholerae* host as measured by qPCR. Ectopic expression was induced 20 minutes prior to phage infection. **e,** Plaque size of listed phage on PLE (-) *V. cholerae*. **p<0.001, *ns* not significant

identifiable homolog found in a *Pseudoalteromonas* phage that is also predicted to have the same DUF2493 C-terminus. These proteins cluster on a more distant branch than the HelA homologs (Supplementary Fig. 2.4B), supporting the hypothesis that *helB* was horizontally acquired by ICP1. Altogether, SF1B helicases are readily found in marine phages, and ICP1 encoding *helA* are the dominant ICP1 shed by cholera patients in Dhaka between 2001-2017.

Most epidemic sampling of ICP1 from cholera patients has been done in Dhaka; however, we recently began sampling cholera patients at a rural site in Mathbaria, Bangladesh. In contrast to what was observed in ICP1 isolates from Dhaka in the 2017 epidemic period, all the ICP1 isolates recovered from cholera patients in Mathbaria encoded the *helB* allele (Fig. 2.4C). One representative isolate from Mathbaria from 2017, referred to here as ICP1^{B}, is over 99.8% identical to ICP1^{A} across 90% of the genome, with 205 of 227 ICP1^{B} predicted open reading frames being shared with ICP1^{A}. The resurgence and dominance of *helB* in the Mathbaria epidemic sampling suggests that there could be a selective advantage for ICP1 encoding *helB* rather than *helA* in this region.

As ICP1^{B} is not isogenic to ICP1^{A}, we first wanted to characterize the role of *helB* in ICP1^{B} fitness. Similar to HelA, HelB is able to be detected by Western blot within 8 minutes of infection (Fig. 2.5A), again coinciding with ICP1 replication [23]. Also similar to *helA*, *helB* is not essential for ICP1^{B}, and ICP1^{B} *ΔhelB* is able to form plaques on both PLE (-) and PLE (+) hosts (Fig. 2.5B). Interestingly, ICP1^{B} *ΔhelB* forms plaques on PLE (+) *V. cholerae* with a higher efficiency than ICP1^{A} *ΔhelA*, suggesting that ICP1^{B} has evolved other ways to limit PLE-mediated anti-phage activity.

We next wanted to see if ICP1^{B} replication was impacted by the *helB* deletion. In contrast to *ΔhelA* in ICP1^{A}, ICP1^{B} *ΔhelB* demonstrates a large drop in ICP1^{B} copy number during the course of infection (Fig. 2.5C), indicating that although *helB* is not necessary for ICP1^{B} replication, it does have a more central role in phage fitness. Consistent with the observation that PLE decreases the ability of ICP1^{A} to replicate (Fig. 2.1E), replication of ICP1^{B}, too, is impacted negatively by PLE; however, ICP1^{B} *ΔhelB* does not restore the ability of ICP1^{B} to replicate in the presence of PLE (Fig. 2.5C), demonstrating a more severe fitness effect associated with losing the accessory helicase on ICP1^{B} than on ICP1^{A} independent of the presence of PLE.

To confirm the role of *helB* in diminished ICP1^{B} fitness, we next ectopically expressed *helB* to complement the mutant phage. ICP1^{B} *ΔhelB* is unable to be complemented by ectopic expression of *helB*, suggesting that the observed decrease in ICP1 fitness may not be due to direct loss of the *helB* gene product (Fig. 2.5D)**.** To minimize potential polar effects of Δ*helB*, a targeted mutation was made to remove 25 amino acids encompassing the helicase domain (HD) that contains the Walker A motif necessary for ATP hydrolysis [77]. While ICP1^{B} *helB ΔHD* had increased phage replication relative to the clean *helB* deletion, there was still a defect in replication that could not be complemented (Fig. 2.5D), suggesting that ectoptic expression may not be able to achieve the appropriate timing or dosage of *helB* expression, or that the fitness cost is not a result of loss of HelB *per se*. Due to the complex nature of phage genomes and tight regulation of phage gene expression, disruption of even the HD domain of *helB* could have detrimental effects on uncharacterized *in cis* sites that could contribute to poor fitness. The fitness defect associated with mutant *helB* was also observed as a decrease in plaque size, with both ICP1^{b} *ΔhelB* and ICP1^{B} *helB ΔHD* forming plaques that are on average less than 0.66 times the size of ICP1^{B} (Fig. 2.5E). Altogether, ICP1^{B} is less fit in the absence of *helB*, consistent with the observation that all natural ICP1 isolates encode an SF1B-type helicase.

PLE exploits phage-encoded SF1B-type helicases to drive replication during ICP1 infection

Given that PLE replication requires *helA* (Fig. 2.2A), and ICP1 with *helB* have resurged in Mathbaria*,* we were tempted by the possibility that phage with *helB* could be selected for as a mechanism to impede PLE replication during infection. Hence, we next assessed if *helB* could also support PLE replication. Consistent with the inverse relationship between ICP1 and PLE replication, PLE still replicated when infected with ICP1^{B}, and PLE replication was not observed in the absence *helB* (Fig. 2.6A), indicating that *helB* is also necessary for PLE replication despite HelB having less than shared 25% shared amino acid identity with HelA (Supplementary Fig. 2.4A). Further, ectopic expression of *helB* complemented the defect in PLE replication observed during infection with ICP1^{B} *ΔhelB,* and ectopic expression of *helA* was likewise sufficient to restore PLE replication during infection with ICP1^{B} *ΔhelB* (Fig. 2.6A). These data demonstrate that PLE is able to harness either helicase independent of the ICP1 isolate that is infecting the host. Additionally, the shared ability of these non-isogenic ICP1

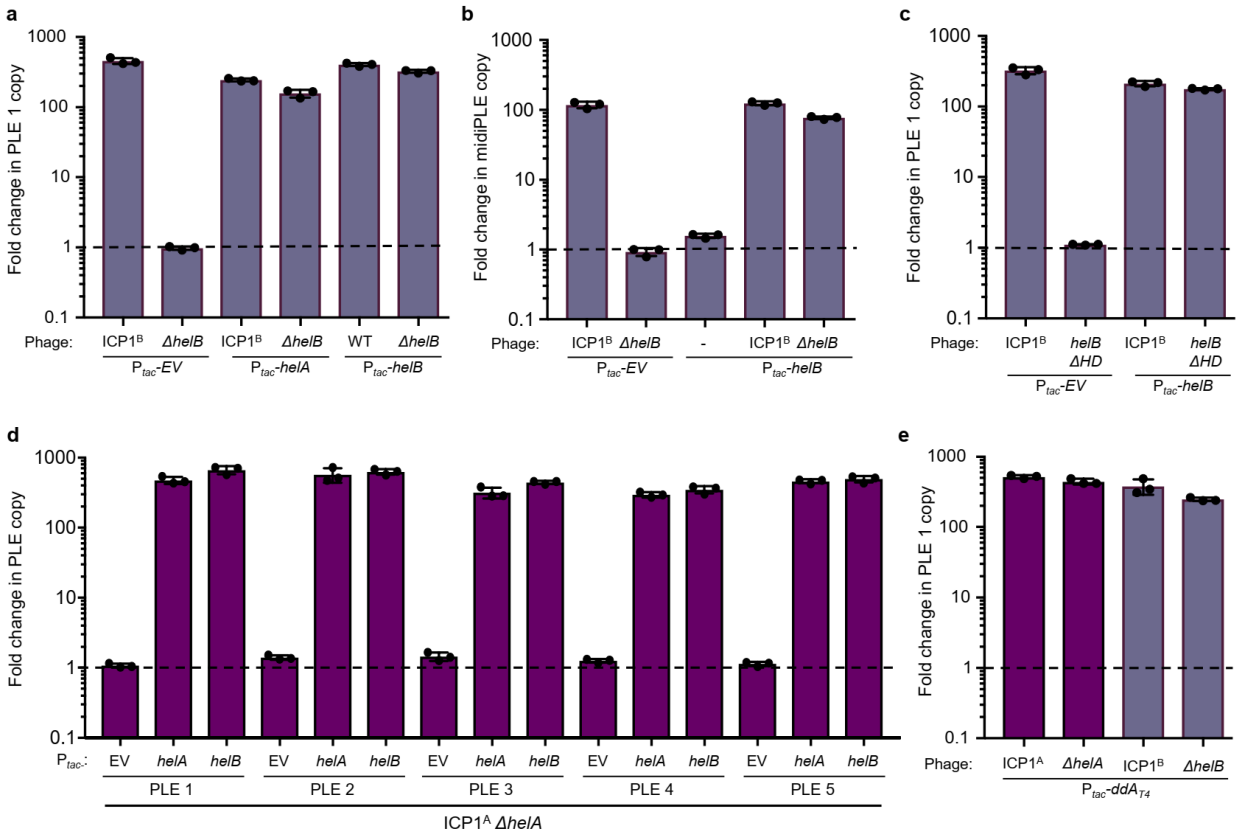


Figure 2.6. ICP1-encoded *helB* is necessary for PLE replication. a, Replication of PLE (A,C) or midiPLE (B) 20 minutes following infection by the listed ICP1^{B} variant as measured by qPCR. Ectopic vectors were induced 20 minutes prior to infection. Dashed line indicates no change in copy. **b,** Replication of the listed PLE in isogenic hosts 20 minutes following infection by ICP1^{A} *ΔhelA*. Ectopic vectors were induced 20 minutes prior to infection. Dashed line indicates no change in copy. **c,** Replication of PLE 20 minutes following infection by the listed phage as measured by qPCR. Ectopic expression of *dda* from *E. coli* phage T4 was induced 20 minutes prior to infection. Dashed line indicates no change in copy.

isolates to facilitate PLE replication also implicates additional, functionally conserved ICP1 gene products, in addition to *helA* and *helB,* in PLE replication.

Similar to *helA,* we next used ICP1^{B} *ΔhelB* to probe for midiPLE replication following ectopic expression of *repA*. As expected, midiPLE is able to replicate when infected with ICP1^{B} but failed to replicate in the absence of *helB*, indicating that *helB* is also directly involved in PLE replication (Fig. 2.6B). Like *helA*, *helB* is also not sufficient to stimulate PLE replication in the absence of ICP1^{B}, showing that PLE is still dependent on additional replication machinery from ICP1^{B}. We additionally confirmed that the ability of HelB to hydrolyze ATP is required for HelB to facilitate PLE replication by testing the ICP1^{B} *helB ΔHD* variant, and, as anticipated, the helicase activity of *helB* is necessary for PLE replication (Fig. 2.6C).

The first ICP1 isolate identified with the *helB* allele was from Dhaka in 2006 when PLE 2 *V. cholerae* was being shed by cholera patients [22,29], leading us to evaluate if the two helicase alleles have different capacities to facilitate replication of different PLEs during infection with ICP1. To test this hypothesis, we first infected isogenic *V. cholerae* harboring each of the five characterized PLEs with ICP1^{A} and observed that all PLEs replicated equally well (Supplementary Fig. 2.6). Next, we determined that *helA* is

necessary for replication of all five PLEs during ICP1^{A} infection and that replication can be complemented with ectopic expression of *helA* (Fig. 2.6D). To evaluate if each PLE can additionally use *helB* to support replication, we also complemented ICP1^{A} *ΔhelA* with ectopic expression of *helB* and found that in fact all five PLEs can use either one of the two ICP1-encoded accessory helicases for replication.

As current data supports the model that PLE responds specifically to ICP1 infection [22,29], we next wanted to determine if PLE's capacity to exploit either *helA* or *helB* to drive PLE replication is specific to ICP1-encoded proteins or to a more general SF1B type helicase. To address the specificity of the interaction, we ectopically expressed the SF1B-type helicase *dda* from *E. coli* phage T4 during infection with either ICP1^{A} *ΔhelA* or ICP1^{B} *ΔhelB*. T4 Dda is only 16% identical to either HelA or HelB, and does not group with the marine phage SF1B type helicases (Supplementary Fig. 2.4B). Although PLE cannot replicate while infected with either of these phage alone (Figures 6A and 6D), expression of *dda* was sufficient to support PLE replication in the absence of ICP1-encoded accessory helicases (Fig. 2.6E). Despite the apparent specificity between PLE and ICP1, the ability of PLE to exploit a variety of phage-encoded accessory helicases reveals flexibility in at least one requirement for PLE replication, and suggests that swapping of helicase alleles by ICP1 isolates is not a beneficial strategy to mitigate PLE parasitism.

Discussion

As a defense island and phage parasite of ICP1, the *V. cholerae* PLE has become highly evolved to make use of phage-encoded gene products to drive its anti-phage program [29] We characterize here a new ICP1-PLE interaction: PLE makes use of a non-essential ICP1-encoded SF1B type helicase, *helA*, to drive PLE replication during infection (Fig. 2.2A). In addition, we see that PLE has evolved to make use of not just two unrelated helicases encoded by ICP1, but also of T4 *dda*, the prototypical but similarly unrelated SF1B helicase (Figures 4B and 4D), implicating strong evolutionary pressures for maintenance of PLE replication in response to ICP1 infection. PLE replication is necessary for optimum gene expression (Fig. 2.2C) as well as for PLE transduction [23], highlighting the beneficial roles that *helA* plays for PLE. However, the rates of PLE transduction are very low per cell relative to the high copy number that PLE reaches during ICP1 infection [23], suggesting an additional role for PLE replication in ICP1 inhibition.

In comparison to the well-studied phage parasites, SaPIs, PLE is dependent upon its helper phage for replication directly as evidenced by the requirement for *helA* for midiPLE replication (Fig. 2.2B). SaPIs, on the other hand, make use of their bacterial host's replication machinery and are able to autonomously replicate in the absence of helper phage induction [6]. The unique requirement for the phage-encoded helicase also underscores the differences between the helper phages that induce these chromosomal islands, with PLE being induced by a lytic phage that encodes its own replication machinery and SaPIs by an activated lysogen that also exploits its host-encoded replication machinery [6]. The fact that *helA* expression alone is not sufficient to drive midiPLE replication in the absence of ICP1 infection implicates other ICP1-encoded replication proteins in facilitating PLE replication or *V. cholerae* encoded proteins that are only expressed during infection (Fig. 2.2B). Aside from the SFIB type helicases, the

potential role for ICP1's replication machinery in PLE replication remains to be elucidated. As T4 *dda* has been observed to have a role in T4 origin initiation during origin-dependent replication [75], we speculate that *helA* has a similar role in facilitating origin firing in PLE by interacting with a conserved part of the PLE machinery and recruiting a conserved ICP1 replication protein. ICP1 is predicted to encode a DNA polymerase and primase/helicase reminiscent of machinery that drives the *E. coli* phage T7 replisome [78]. Further work remains to identify what roles, if any, these replisome proteins have in PLE mobilization.

Despite not being essential (Figures 1B and 1E, Figures 5B and 5C), all ICP1 isolates encode an accessory SF1B type helicase, as do several marine phages (Supplementary Fig. 2.4B), suggesting that ICP1 could be exchanging genetic material with or could be related to these marine vibriophages. The fitness costs, as measured by plaque size, implicate both *helA* and *helB* in maintaining a healthy phage population (Figures 4A and 5E). The ease with which PLE is able to make use of ectopically expressed *helB* compared to the inability of ectopically expressed *helB* to complement the ICP1^{B} *ΔhelB* replication deficiency suggests that these helicases play a specialized role in the phage lifecycle that is more complex than for PLE.

The striking spatial separation between the ICP1^{A} and ICP1^{B} populations that were shed by cholera patients during the same epidemic period (Fig. 2.4C) suggests that slight variations in the phage strain, such as the difference between *helA* and *helB*, can have large differences in the makeup of phage populations. Indeed, the ability of ICP1^{B} *ΔhelB* to form plaques in the presence of PLE (Fig. 2.5B), as well as the ability of PLE to select for ICP1^{B} with mutations in *helB* suggests that ICP1^{B} should dominate in the presence of PLE (+) *V. cholerae* (Supplementary Fig. 2.5A); however, the greater fitness cost in the absence of *helB*, as evidenced by the diminished ability of ICP1^{B} *ΔhelB* to replicate, suggests that the loss of *helB* in the presence of PLE (+) *V. cholerae* is ultimately detrimental to the phage population. Due to the lack of sampling history in Mathbaria where ICP1^{B} was isolated, no models can be made to examine the role of *helB* on PLE prevalence in patient samples, although it is enticing to speculate that ICP1^{B} is more fit to flourish in the absence of PLE where it has less of a chance of mutating, while ICP1^{A} is more fit to combat PLE due to the milder *ΔhelA* phenotype, which is dominant in Dhaka patient samples where PLE is highly shed [69].

In order to defend against viral infection, host resistance mechanisms must have ways by which they bypass virus mediated host takeover. Eukaryotic DNA and RNA viruses broadly use virally encoded ribonucleases to globally degrade host transcripts in the infected cell, which sabotage their hosts through modulation of transcript and protein levels. This decrease in transcript abundance leads to a downregulation of innate immune responses, processes which are detrimental to the host but are ultimately reversable [79,80]. Conversely, degradation of the host chromosome is a host takeover process that is unique to phages. Host chromosome degradation has a twofold benefit for the predatory phage: it cleaves and releases nucleosides that can then be incorporated into the rapidly replicating phage genome, but it can also destroy the template of anti-phage genes encoded by the bacterial host. Consistently, when PLE is unable to replicate or excise from the chromosome, it is thus unable to block plaque formation by ICP1 (Fig. 2.1B) and is susceptible to ICP1-encoded DNases (Fig. 2.3).

The necessity of excision and replication of PLE during ICP1 infection highlights the crucial role that mobilization of inducible phage defense systems has during phage infection. In order for inducible defenses to functionally protect a host cell from phage infection, they must be able to overcome the infecting phage's destruction of the host chromosome. It thus stands to reason that the high probability of phage defense systems being encoded on genomic islands may be in part due to the ability of genomic islands to mobilize during infection and escape phage-mediated host takeover, with the potential of horizontal transfer as an added benefit.

Methods

Bacterial growth conditions

The bacterial strains, plasmids, primers and phages used in this study are listed in Supplementary Tables 2.2-2.5. All bacterial strains were grown at 37 °C in LB with aeration or on LB agar plates. The following antibiotics were used as necessary: streptomycin (100 μg/mL), spectinomycin (100 μg/mL), kanamycin (75 μg/mL), ampicillin, (*V. cholerae* 50 μg/mL, *E. coli* 100 μg/mL), chloramphenicol (*V. cholerae* 1.25 μg/mL, *E. coli* 25 μg/mL). Ectopic expression constructs in *V. cholerae* were induced 20 minutes prior to ICP1 infection with 1 mM IPTG and 1.5 mM theophylline. Mutants were cloned using SOE (splicing by overlap extension) PCR and introduced by natural transformation [50]. Plasmids were constructed using Gibson Assembly or Golden Gate Assembly (New England Biolabs).

Phage conditions

Phage were propagated using the soft agar overly method and high titer stocks were made by polyethylene glycerol precipitation [49]. Total phage gDNA was prepped with a DNeasy Blood & Tissue Kit (Qiagen). Phage mutants were constructed using CRISPR-Cas engineering as previously described [29,53]. Spot plates were performed as before [29] but with slight modification. Briefly, mid-log *V. cholerae* was added to 0.5% molten LB agar that was added to a solid agar plate and allowed to solidify. Ten-fold dilutions of phage were applied to the surface in 3 μL spots and allowed to dry. Plates were incubated at 37 °C. Images are representative of at least two independent experiments. The efficiency of plaquing was calculated by comparing the number of plaques a given phage forms on PLE (-) *V. cholerae* relative to the number of plaques formed on PLE (+) *V. cholerae*. Each EOP is calculated in triplicate, and the limit of detection is the point at the phage is unable to productively infect the PLE (+) host while still forming plaques on a PLE (-) host. Plaque size was determined by imaging at least 20 plaques each from 3 independent replicates in 0.5% agar overlay on PLE (-) *V. cholerae*.

qPCR conditions

Fold change in genome copy was performed as before [22] but with slight modification. For the fold change in ICP1 copy number, cultures were grown to an OD = 0.3, infected with a multiplicity of infection (MOI) of 0.01. Samples were taken upon ICP1 addition and 20 minutes following infection, boiled, and diluted 1:50 to be used as template. Quantification of fold change in PLE copy number was achieved from an MOI of 2.5

infection, with samples taken immediately before ICP1 infection and 20 minutes following infection. Boiled samples were diluted 1:1000 and used as template. Quantification of fold change in miniPLE copy number was measured from an MOI of 5 infection, with samples taken immediately before ICP1 infection and 30 minutes following infection. Boiled samples were diluted 1:100 to be used as template. For complementation experiments, ectopic expression was induced 20 minutes prior to infection.

Western Blots

PLE (-) *V. cholerae* was grown to an OD = 0.3 and infected with the endogenously FLAG-tagged ICP1 listed at an MOI of 5. At the listed timepoints, 1 mL samples were collected and mixed with equal volume ice-cold methanol and centrifuged at 13000 rpm for 3 minutes at 4 °C. Pellets were washed with ice-cold PBS, resuspended in 1x Laemmli buffer, and boiled for 10 minutes at 99 °C. Total protein was run on a 10% SDS-PAGE gel. Primary Rabbit-α-FLAG antibodies (Sigma) were used at a dilution of 1:5000 and detected with goat-α-rabbit-HRP conjugated secondary antibodies at a dilution of 1:5000 (Bio-rad). Clarity Western ECL Substrate (Bio-rad) was used to develop the blots and a Chemidoc XRS Imaging System (Bio-rad) was used to image. Images are representative of at least two independent experiments.

Lysis kinetics and Nanoluciferase assay

PLE (+) *V. cholerae* cells were grown to an OD = 0.2 and induced with IPTG and theophylline for 20 minutes. Cells were then normalized to an OD = 0.3 and infected at an MOI of 2.5. For lysis kinetics, OD_{600} was monitored for 30 minutes. For nanoluciferase, 100 μL cells were sampled at T=0 and T=20 and added to 100 μL ice cold methanol. Luminescence was measured in a Spectra Max i3x plate reader (Molecular Devices) using the Nano-Glo Luciferase Assay System (Promega). Relative luminescence was calculated by dividing the luminescence of the knockout phage relative to the luminescence of the WT phage.

PCR conditions

Plaque circularization PCRs was performed as described [29]. Detection of the *helA* and *helB* alleles from performed on 1 μL ICP1 gDNA prepped as above. PCRs were run on 2% agarose gels and imaged with Gel Green.

Computational analysis

Escape ICP1 *ΔpexA* phage were isolated from and purified twice on PLE (+) *V. cholerae*. Total gDNA was prepped as above. NEBNext Ultra II DNA Library Preparation Kit for Illuminia (New England Biolabs) was used to prep genomic DNA and was sequenced by paired-end sequencing (2 × 150 bp) on an Illumina HiSeq4000 (University of California, Berkeley QB3 Core Facility). The wild-type phage genome was assembled using SPAdes [81] with paired-end reads and default settings. This assembly was used as the reference sequence for comparison to escape phage sequence reads with breseq [82] in 'consensus' mode and default settings. Protein alignments were analyzed using Praline [58]. HelA conservation was determined by analyzing HelA from 25 phages isolated between 2001 and 2017 (Supplementary Table 2.6) [19,25,69]. Phages

included in the phylogenetic analysis were selected from a BLASTP search of HelA and HelB. Each hit was included if it had over 30% identity to either protein across 90% of the protein. A multiple alignment of helicase amino acid sequences was generated with MUSCLE v3.8.31 [83] using default settings. The alignment file was converted to the PHYLIP format with Clustal X v2.0 [84] and a bootstrapped (n=100) maximum-likelihood phylogenetic tree was solved using PhyML v20120412 [85] with the following settings: -d aa -s BEST --rand_start --n_rand_starts 100 -o tlr -b 100).

Chapter 4

Competition between mobile genetic elements drives optimization of a phage-encoded CRISPR-Cas system: Insights from a natural arms-race[69]

Amelia C. McKitterick, Kristen N. LeGault, Angus Angermeyer, Munirul Alam, and Kimberley D. Seed

Summary

CRISPR-Cas systems function as adaptive immune systems by acquiring nucleotide sequences called spacers that mediate sequence-specific defense against competitors. Uniquely, the phage ICP1 encodes a Type I-F CRISPR-Cas system that is deployed to target and overcome PLE, a mobile genetic element with anti-phage activity in *Vibrio cholerae*. Here, we exploit the arms race between ICP1 and PLE to examine spacer acquisition and interference under laboratory conditions to reconcile findings from wild populations. Natural ICP1 isolates encode multiple spacers directed against PLE, but we find that single spacers do not equally interfere with PLE mobilization. High-throughput sequencing to assay spacer acquisition reveals that ICP1 can also acquire spacers that target the *V. cholerae* chromosome. We find that targeting the *V. cholerae* chromosome proximal to PLE is sufficient to block PLE and is dependent on Cas2-3 helicase activity. We propose a model in which indirect chromosomal spacers are able to circumvent PLE by Cas2-3-mediated processive degradation of the *V. cholerae* chromosome before PLE mobilization. Generally, laboratory acquired spacers are much more diverse than the subset of spacers maintained by ICP1 in nature, showing how evolutionary pressures can constrain CRISPR-Cas targeting in ways that are often not appreciated through *in vitro* analyses.

Introduction

Phages often vastly outnumber their bacterial hosts in a variety of environments [86]. As such, bacteria have evolved numerous mechanisms for phage defense, including adaptive immunity via clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins [26,30]. CRISPR-Cas systems are composed of a CRISPR array—a series of "spacers" of foreign sequence alternating with repeats that are transcribed into CRISPR RNAs (crRNAs)—and CRISPR-associated (Cas) genes. Together with crRNAs, Cas proteins defend against foreign nucleic acids, such as the genome of an infecting phage, through a three-step process: adaptation, crRNA and *cas* gene expression, and interference. During adaptation, a foreign DNA fragment is incorporated into the CRISPR array to provide a molecular memory of the challenges that the host cell has faced. This CRISPR array is expressed and processed into individual crRNAs, which complex with Cas proteins and survey the cell for complementary invading nucleotides. Upon finding a complementary sequence, termed protospacer, a Cas nuclease is recruited to the site to mediate interference by cleaving the substrate, ultimately leading to the destruction of the invader [26,87]. Across CRISPR-Cas containing bacteria and archaea, Class 1 Type I CRISPR-Cas systems employing a Cas3 enzyme for DNA unwinding and degradation [88], are the most prevalent [89].

CRISPR-Cas systems do not discriminate between horizontally acquired traits based on fitness gain or loss. Hence, CRISPR-Cas systems are equally capable of halting harmful invading phage DNA as they are halting beneficial mobile genetic elements, including those encoding antibiotic resistance and pathogenicity genes [90–92]. As such, some pathogens only have alternative anti-phage defense systems [93]. For example, the currently circulating biotype of epidemic *Vibrio cholerae*, the causative agent of the diarrheal disease cholera, does not rely on CRISPR-Cas for phage defense [53]. Instead, *V. cholerae* evolved to use phage inducible chromosomal island-like elements (PLEs) to defend against the prevalent lytic phage, ICP1 [22]. PLEs are mobile genetic elements that reside integrated in the small chromosome of *V. cholerae* [22]. During ICP1 infection of PLE (+) *V. cholerae*, PLE excises from the host chromosome, replicates to high copy and is horizontally transduced to naïve neighboring cells, all the while inhibiting phage replication through unknown mechanisms (Fig. 3.1a).

In order to overcome the anti-phage activity encoded by *V. cholerae* PLE, some ICP1 isolates use a Type I-F CRISPR-Cas system that directly targets PLE (Fig. 3.1a), making the CRISPR-Cas system essential for the phage to form plaques on PLE (+) *V. cholerae* [24]. Type I-F systems are composed of three Csy proteins that make up the Csy complex along with Cas6f, a protein involved in crRNA processing [94]. This complex interacts with the processed crRNA to search DNA for a complementary protospacer with an appropriate self versus non-self discrimination sequence, known as the protospacer adjacent motif (PAM) [95]. Upon finding a match with an appropriate PAM, the trans-acting Cas2-3 fusion protein is recruited to degrade the target DNA. In addition to endonuclease activity, Cas2-3 has a helicase domain that unwinds DNA as the protein translocates away from the target DNA, allowing for continued processive degradation of adjacent DNA *in vitro* [96,97]. Recently, sequence analysis identified phages that are predicted to encode CRISPR arrays and/or Cas genes [98,99]; however, ICP1 is the only phage shown to encode a fully functional CRISPR-Cas system [22,24].

As is true when CRISPR-Cas is harnessed by a prokaryotic host for genome defense, the ICP1-encoded CRISPR-Cas system is tasked with targeting and degrading a hostile mobile genetic element. However, there are additional challenges associated with a phage encoding and relying on CRISPR-Cas for its own survival. The ICP1 infection cycle occurs over a 20 minute period, and current data suggest that ICP1 synthesizes its CRISPR-Cas machinery *de novo* upon infection of *V. cholerae* [24]. PLE is induced to excise within minutes of infection through interactions with an early phage-encoded gene product [29]. Thus, in order to overcome PLE, CRISPR synthesis and interference must outpace a rapidly replicating target.

ICP1 and *V. cholerae* are consistently co-isolated from patient stool samples in regions where cholera is endemic such as Bangladesh [19,22,25]. Five genetically distinct PLE variants in *V. cholerae* have appeared in temporally discrete waves across cholera epidemics [22]. Previous analysis revealed that ICP1-encoded CRISPR-Cas can adapt and acquire new spacers against PLE under laboratory conditions [24], however the rules governing spacer acquisition and targeting efficacy for this system are not known. Further, recent comparative genomics of 18 ICP1 isolates collected from Bangladesh between 2001-2012 found that 50% carry CRISPR-Cas [21], however the contemporary state of circulating ICP1 and *V. cholerae* PLE in the region are not known.

Here, we provide an up-to-date understanding of the genomic variants of ICP1 and PLE circulating in Bangladesh. We find that natural ICP1 isolates encode multiple anti-PLE spacers and experimentally validate that increased PLE targeting by ICP1 is required to fully abolish PLE mobilization. Significantly, using a high-throughput spacer acquisition assay and experimental validation, we show that noncanonical PAMs and indirect protospacers in the *V. cholerae* small chromosome can unexpectedly provide protection against PLE. Our results support a model in which ICP1-encoded CRISPR-Cas that is directed against the *V. cholerae* small chromosome is in a race to reach PLE before it excises from the chromosome to exert its anti-phage activity. Taken together, our study highlights the differences between interference competent spacers under laboratory conditions and those that are selected for in nature to provide mechanistic insight into the evolutionary pressures governing the interactions between epidemic *V. cholerae* and its longstanding battle with the predatory phage ICP1.

Results

ICP1-encoded CRISPR-Cas is fixed in the natural phage population

We set out to compare ICP1 and PLE from contemporary cholera patient stool samples to previously identified isolates from the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) in Dhaka, Bangladesh [24,25]. We isolated eight new ICP1 isolates from cholera patient stool samples collected between 2015-2017 and found that all isolates harbor CRISPR-Cas. Thus it appears that ICP1 isolates lacking CRISPR have not been identified in Bangladesh since 2006 [21]. Analysis of the CRISPR arrays indicates a strong selection for spacers specifically targeting PLE (Fig. 3.1b, Supplementary Table 3.5), supporting the function of the ICP1-encoded CRISPR-Cas system as a counter-attack against the anti-phage island PLE [24]. To evaluate if the fixation of CRISPR in ICP1 is necessitated by co-circulating PLE in epidemic *V. cholerae*, we determined the prevalence of PLE over the same near two-decade long period in Dhaka. Combined with previous analyses [22,25], we observed an increase in the

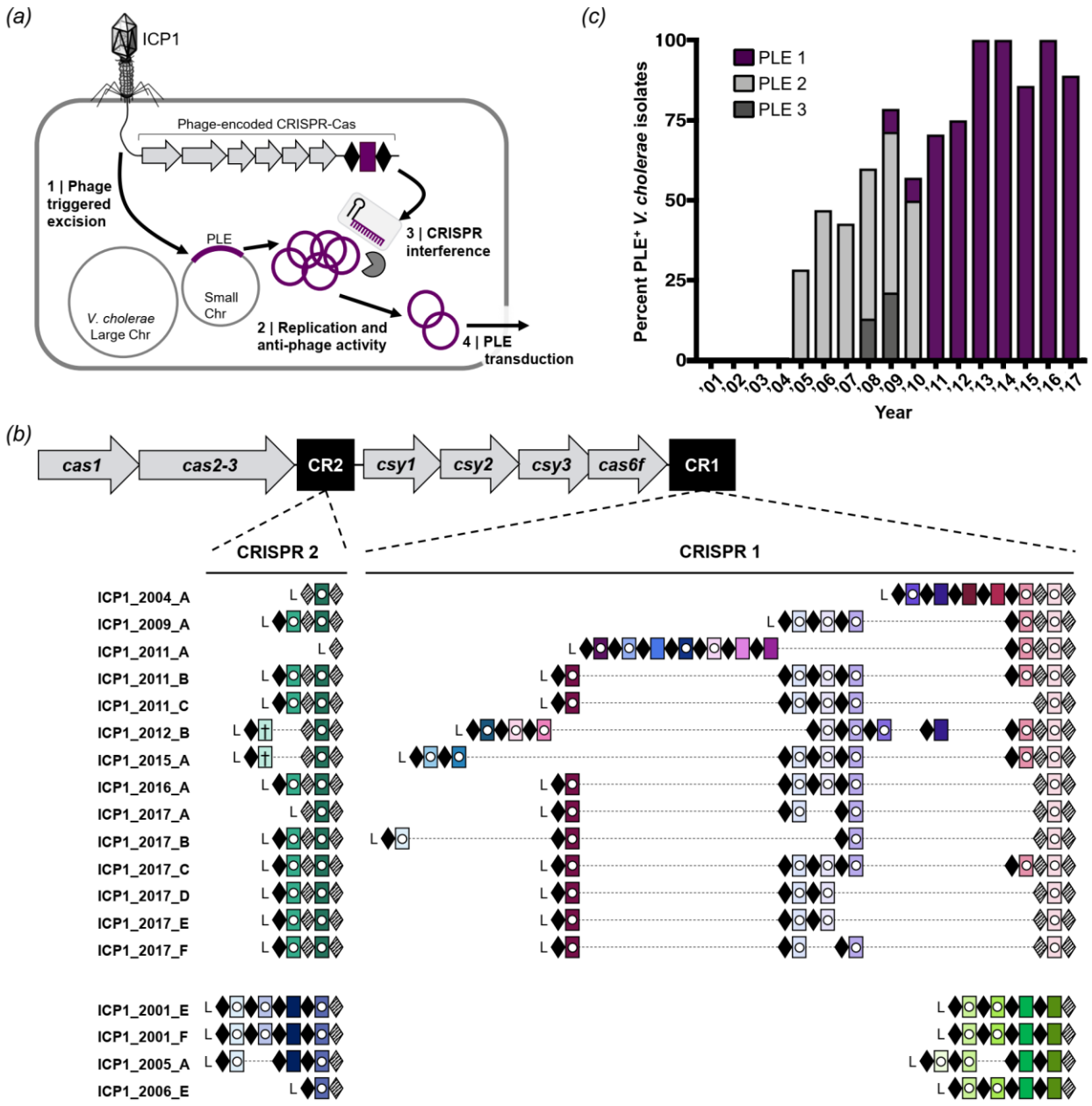


Figure 3.1. ICP1 uses CRISPR-Cas to overcome epidemic *V. cholerae* PLE. a, Lytic phage ICP1 infects *V. cholerae* triggering PLE excision. PLE replicates and exerts anti-phage activity, ultimately leading to PLE transduction. Concurrently, ICP1-encoded CRISPR-Cas is expressed to interfere with PLE activity. **b,** The architecture of the ICP1 CRISPR-Cas system and comparison of spacer composition between phage isolates. For each CRISPR locus, the repeat (28 bp) and spacer (32 bp) content is detailed as black diamonds and colored rectangles, respectively. Repeats that match the repeat consensus [24] are shown in solid diamonds, and degenerate repeats are indicated in hatched black diamonds. An AT-rich leader sequence (L) precedes each CRISPR locus. Identical spacers shared between isolates are shown as rectangles with identical colors. Spacers containing a white circle target PLE, and spacers containing a cross target the *V. cholerae* large chromosome. **c,** Percentage of *V. cholerae* isolates harboring PLE recovered from epidemic sampling at the ICDDR,B over time (n=230 strains analyzed).

prevalence of PLE (+) *V. cholerae* in epidemic sampling over time (Fig. 3.1c). Of note is the high prevalence of PLE 1 *V. cholerae* over the past 6 years, indicating that this variant of the anti-phage island is currently dominating the epidemic landscape in Dhaka. Despite the relatively long period over which PLE 1 has been dominant in Dhaka, and consistent with previous results[22,25], whole genome sequencing of eight PLE 1 *V. cholerae* isolates showed that PLE 1 is 100% identical at the nucleotide level in all strains.

Multiple spacers increase ICP1 CRISPR-Cas mediated PLE interference

All of the natural phage we isolated encode multiple CRISPR spacers against PLE (Fig. 3.1b); however, previous work revealed that only one functional spacer is required for ICP1 to overcome PLE-mediated anti-phage activity as evaluated by plaque formation (Supplementary Fig. 3.1) [24]. Conversely, a single spacer against the PLE did not prevent transduction of PLE [22]. To investigate the consequences of varying spacer number and identity on PLE transduction and replication, we used co-isolated ICP1 and PLE 1 *V. cholerae* obtained from a cholera patient sample in 2011 [24]. This ICP1 isolate harbors two spacers (spacers 8 and 9) at the leading edge of the CRISPR 1 array that target PLE 1. We also used an isogenic phage with a spontaneous loss of spacer 9 [24], as well as one that acquired an additional 10th spacer targeting PLE *in vitro* (Fig. 3.2a). Despite the ability to overcome PLE and form plaques, spacer 8 targeting was not sufficient to decrease PLE transduction during ICP1 infection relative to an untargeted control (Fig. 3.2b). In comparison, two anti-PLE spacers decreased PLE transduction during ICP1 infection and three spacers completely abolished PLE transduction, showing that increased CRISPR targeting by ICP1 has a stronger anti-PLE effect. To evaluate potential differences between spacer 8 and spacer 9 on

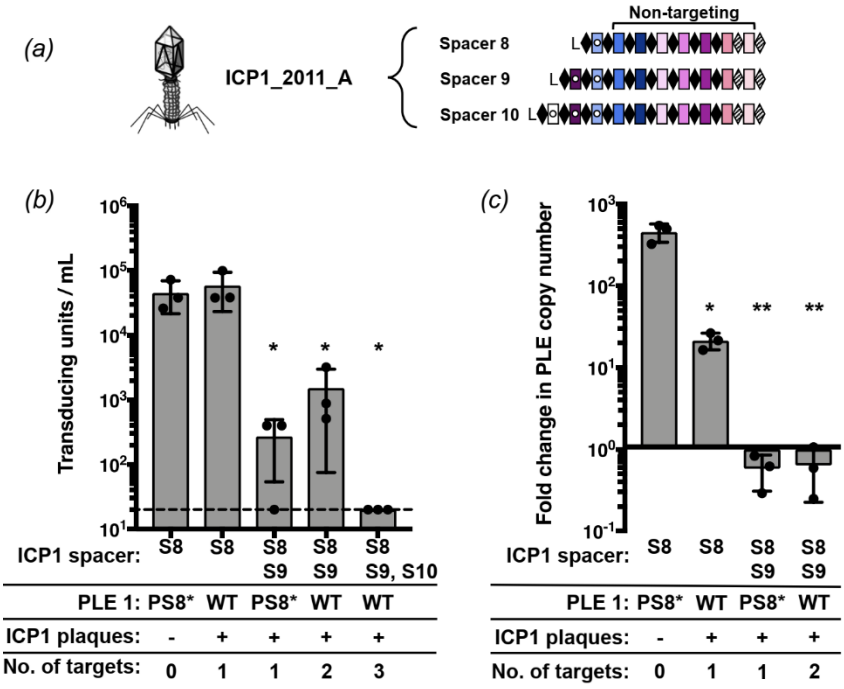


Figure 3.2. CRISPR can limit horizontal transmission of PLE. a, ICP1_2011_A with anti-PLE 1 spacers S8, S9 and S10 (shown with internal white circles) tested in panels **b** and **c**. **b,** PLE transduction after infection with ICP1 with 0,1,2 or 3 spacers. The dashed line indicates the limit of detection for this assay. PS8* indicates silent mutations in protospacer 8 that abolishes CRISPR interference [24]. A single spacer is necessary and sufficient to permit lytic growth of ICP1 on PLE 1 *V. cholerae* as seen by equal plaque formation (indicated by +, Supplementary Fig. 3.1). **c,** PLE replication 20 minutes after infection with ICP1 with 0,1 or 2 spacers as determined by qPCR. For panels b and c, error bars indicate standard deviations of biological triplicates. Significance was determined by T Test, * $p<0.05$, ** $p<0.005$.

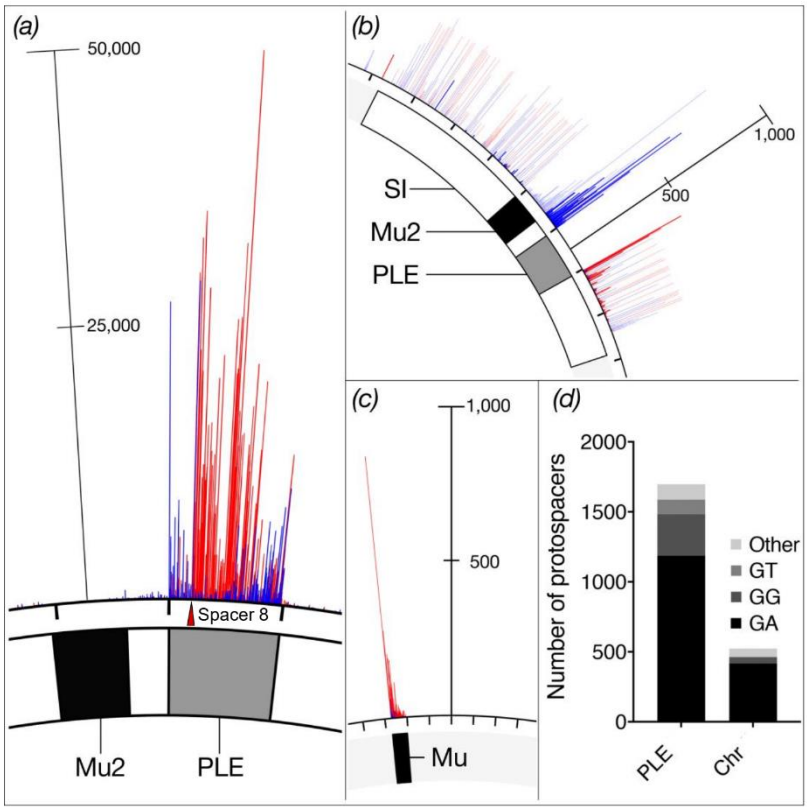


Figure 3.3. High-throughput interference driven spacer acquisition mapping. a, The locations of the ICP1 CRISPR leader-proximal spacer on the *V. cholerae* small chromosome. The location of the interference-efficient spacer (S8) is indicated with the red triangle. **b,** Spacer locations on the *V. cholerae* small chromosome (PLE mappings not shown for clarity). Uniquely mapped spacers are shown in solid blue or red, while translucent bars show mapping of spacers to all possible locations. **c,** Spacer locations on the *V. cholerae* large chromosome. For panels a, b and c, spacers on the plus and minus strand are indicated in red and blue, respectively. The scale bar measures the number of mapped spacers, and the tick marks around the chromosome are in 18kb intervals. The white box represents the superintegron (SI), the black box is the mu-like region and the grey box is PLE 1. **d,** Proportion of unique protospacers with a GA or other dinucleotide PAM sequence in PLE or in the small chromosome (Chr).

PLE targeting, we used PLE 1 with a protospacer mutation (PLE 1^{PS8*}) that inhibits spacer 8-mediated PLE targeting [24]. Strikingly, just spacer 9 targeting PLE alone was able to decrease PLE transduction to the same level as when two spacers were targeting PLE (Fig. 3.2b).

We next analyzed the copy number of PLE during infection with ICP1 encoding one or two targeting spacers to identify if the differences in reducing PLE transduction were due to differences in PLE copy number (Fig. 3.2c). In the absence of ICP1 CRISPR targeting, PLE replicates to high copy number, which facilitates horizontal transmission [22]. Targeting with only one spacer was sufficient to significantly decrease PLE replication, and in agreement with the transduction data, spacer 9 had a stronger inhibitory effect on PLE replication than spacer 8. Sequencing of the newly transduced PLEs showed no mutations in the protospacers. Thus, the transduced PLEs did not escape CRISPR targeting through mutation, but instead, the individual spacers possess different and incomplete abilities to fully block PLE mobilization. Altogether, these results demonstrate that not all spacers selected in nature equally interfere with PLE mobilization and that increasing the number of spacers provides enhanced capacity of ICP1 to interfere with PLE.

Interference-driven spacer acquisition in ICP1 reveals indirect targets and non-canonical PAMs

Since spacer composition variability in nature was lower than we expected (Fig. 3.1b), we next set out to experimentally sample the repertoire of spacers that ICP1 can acquire to overcome PLE. Low-throughput experiments previously demonstrated that

ICP1 can acquire new spacers targeting the PLE under laboratory conditions without the need to overexpress *cas* genes [24]. To further analyze the natural process of interference-driven spacer acquisition in this system, we performed high-throughput sequencing of expanded CRISPR arrays of phage selected on PLE 1 *V. cholerae*. We infected PLE 1 *V. cholerae* with ICP1 containing spacer 8 (Fig. 3.2a), and the recovered lysate was probed for ICP1 progeny with newly acquired spacers that allowed for plaque formation on a PLE 1^{PS8*} host. Illumina sequencing of the leader-proximal spacer in CRISPR 1 allowed us to sample over 10^6 acquired spacers in each replicate experiment (Supplementary Fig. 3.2, Supplementary Table 3.6). In order to accurately map the spacers to the PLE 1 *V. cholerae* host, we performed complete whole-genome sequencing and assembly of the bacterial genome. As was previously reported [22], we found that PLE 1 was integrated in a *V. cholerae* repeat (VCR), of which over 100 repeats intersperse the *V. cholerae* small chromosome in a gene-capture region, the superintegron [100]. In total, 96% of the acquired spacers mapped to PLE (Supplementary Fig. 3.3), while, interestingly, the other 4% mapped to *V. cholerae* chromosomes (Supplementary Table 3.6).

Mapping of the spacers to the small chromosome showed a pattern of strand bias that reflected previous observations in primed acquisition experiments performed in other Type I-F systems [101], with a distribution of acquired spacers 5' of the protospacer on the non-targeted strand and 3' of the protospacer on the targeted strand (Fig. 3.3a). The distribution of spacers acquired 5' of the protospacer on the nontargeted strand were split between the small chromosomal region proximal to the PLE 1 integration site (Fig. 3.3b, Supplementary Fig. 3.4), as well as the 3' end of PLE. Acquired spacers mapping to the *V. cholerae* chromosome were not evenly distributed between the large and small chromosome, but instead ~90% of the chromosomal spacers mapped to the small chromosome (Fig. 3.3b, Supplementary Table 3.6). Spacers that mapped to the large chromosome were restricted to a mu-like region (Fig. 3.3c), which was duplicated in this strain and was also in the small chromosome proximal to PLE (Fig. 3.3b). Acquired spacers mapped uniformly throughout the superintegron, however, this is likely an artifact as the superintegron is highly repetitive. When considering spacers that map to a single site in the small chromosome, we observed an obvious bias for acquired spacers mapping closer to the PLE integration site (Fig. 3.3b, Supplementary Fig. 3.4).

Consistent with $CRISPR^+$ ICP1 isolates from nature (Supplementary Fig. 3.5), the majority (~70%) of the spacers acquired experimentally targeted protospacers in PLE 1 that were flanked by a 3' GA PAM (Fig. 3.3d). Variations from the canonical GA motif would be expected to abolish CRISPR interference. However, ~30% of protospacers in PLE had non-canonical PAMs, and of those, the majority were GG or GT. Previous CRISPR acquisition studies in Type I-F systems indicate that alternative PAMs can be explained by a "slippage" event [101,102]. To identify putative slippage events, we analyzed the sequences adjacent to GG PAMs and found that 45% of GG PAMs have a canonical GA within 3 nucleotides of the PAM position, suggesting that the ICP1 acquisition machinery has a propensity to slip (Fig. 3.4a).

We next wanted to determine if these non-canonical PAMs are functional for PLE interference. Since only the newest spacer was sequenced in our high-throughput assay, we could not rule out that multiple spacers were not acquired within the

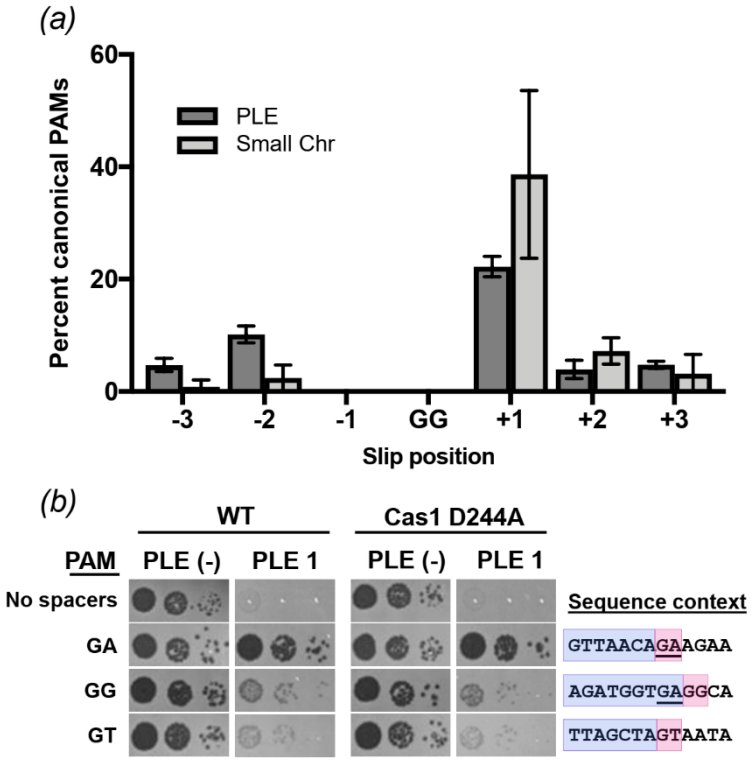


Figure 3.4. Characterizing non-canonical PAMs. a, The frequency of a canonical GA PAM +/- 3nt from a non-canonical GG PAM across all data sets. **b,** Tenfold dilutions of ICP1 engineered to contain a spacer that targets PLE 1 with a non-canonical PAM spotted on *V. cholerae* PLE (-) or PLE 1 lawns showing the ability of different phage strains to form plaques (dark spots, zones of killing) (left). Sequence context (right) of the region adjacent to the PAM. The protospacer is boxed in purple and PAM is boxed in pink. The consensus canonical PAM GA is bolded and underlined.

expanded phage CRISPR array. We therefore engineered ICP1 to encode a single spacer reflective of an experimentally acquired spacer with either the canonical PAM or the most common non-canonical PAMs: either a GG or GT (Fig. 3.3d) and evaluated plaque formation of the engineered phage on PLE 1 *V. cholerae*. Despite relying on a non-canonical PAM, we found that ICP1 is able to target those protospacers and overcome PLE, albeit at a lower efficiency than when targeting a protospacer with a canonical GA PAM (Fig. 3.4b). Even when no canonical PAM was within +/- 3 nt, ICP1 was still able to overcome PLE targeting a protospacer with a GT PAM. As PAM mutations are frequently a source for primed acquisition [103], we tested if the observed residual CRISPR activity was due to further spacer acquisition and interference. We constructed a Cas1 D244A mutation, which disrupts a conserved metal coordinating residue to inhibit spacer acquisition [102] (Supplementary Fig. 3.6) and tested if plaque formation was altered (Fig. 3.4b). We observed no difference in the efficiency of plaque formation between the Cas1 mutants and the parental phage, suggesting that the ICP1 CRISPR-Cas system is more tolerant of divergent PAMs during infection than previously characterized [24].

Protospacers in the small chromosome facilitate ICP1 CRISPR-Cas-mediated PLE interference

In our spacer acquisition experiment, we identified a subset of spacers that target a mu-like region in the *V. cholerae* large chromosome (Fig. 3.3c), suggesting that CRISPR targeting of the mu-like region was advantageous in overcoming PLE. To test the role of protospacers in the mu-like region in PLE interference, we isolated ICP1 that had acquired a spacer that targets the mu-like region and was able to form plaques on PLE 1^{PS8*} (Fig. 3.5). Since assembly of the *V. cholerae* genome revealed that the mu-like region was present and 100% identical in both chromosomes, presumably due to a duplication of the region on the large chromosome (Fig. 3.5a), we wanted to evaluate if targeting the mu-like region *per se* allowed for plaque formation, or if the chromosomal

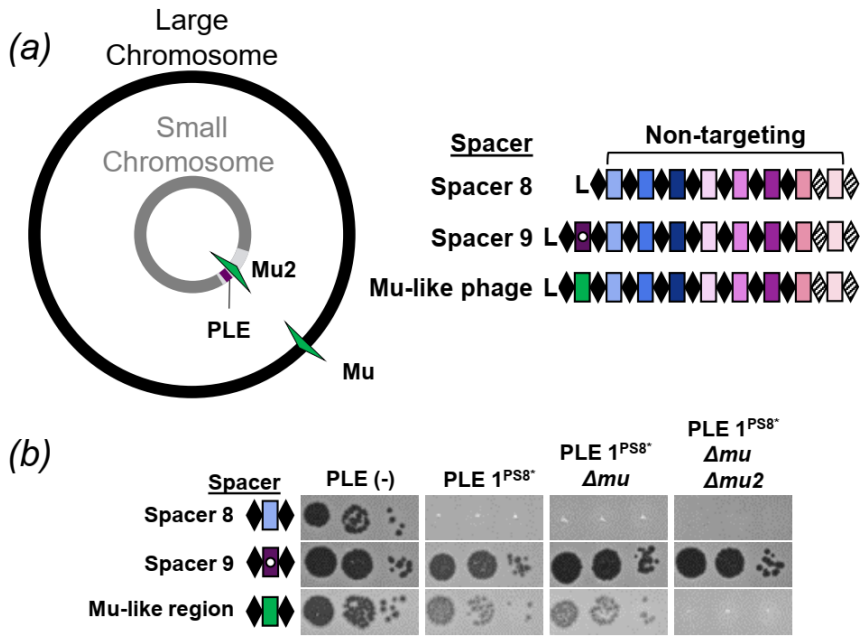


Figure 3.5. ICP1 CRISPR-targeting of the small chromosome facilitates PLE interference. a, Cartoon (left) of the *V. cholerae* large and small chromosomes. The superintegron is shown in light grey, the PLE is shown in purple. The two mu-like regions in the large and small chromosome are shown in green arrows. ICP1_2011_A CRISPR variants (right) used to test the role of targeting sites. The internal white circle indicates the PLE 1 targeting spacer. **b,** Tenfold dilutions of ICP1 with the spacers indicated spotted on *V. cholerae* lawns showing the ability of different phage strains to form plaques.

context was important in allowing for CRISPR-meditated interference with PLE. To test this difference, we generated a single knockout of the mu-like region in the large chromosome and a double knockout in both chromosomes. ICP1 CRISPR-mediated interference with PLE was abolished in the double knockout, however, knocking out the mu-like region in the large chromosome had no effect on ICP1 plaque formation, demonstrating that targeting of the region in the large chromosome is an artifact of the duplication in the small chromosome (Fig. 3.5b). These results show that CRISPR targeting of the *V. cholerae* large chromosome is dispensable for phage overcoming PLE, while targeting the small chromosome is sufficient to overcome PLE activity.

When CRISPR goes off target: going the distance to maintain interference

As processivity of Cas2-3 has been demonstrated *in vitro* [97], we speculated that ICP1 targeting of the small chromosome proximal to PLE interferes with PLE anti-phage activity by the processive degradation of PLE along with the chromosome; however, PLE excises from the chromosome early during ICP1 infection [29]. This timing suggests that CRISPR targeting and Cas2-3 processive degradation of the small chromosome would have to happen prior to PLE excision and would therefore likely be distance dependent. In support of this hypothesis, experimentally acquired spacers mapping to the small chromosomal clustered proximal to PLE (Fig. 3.3b). To test the impact of targeting at increasing distances from PLE, we engineered ICP1 to possess CRISPR arrays containing only one spacer drawn from the experimental acquisition pool that targets the small chromosome at varying distances away from PLE. We then assayed the ability of these engineered phage to overcome PLE and form plaques (Fig. 3.6a). As a positive control, ICP1 engineered with a spacer that targets internal to PLE formed robust and equal plaques on PLE (-) and PLE 1 hosts. In comparison, phage with a spacer that targets far (>400 kb) from PLE were unable to form plaques on PLE 1.

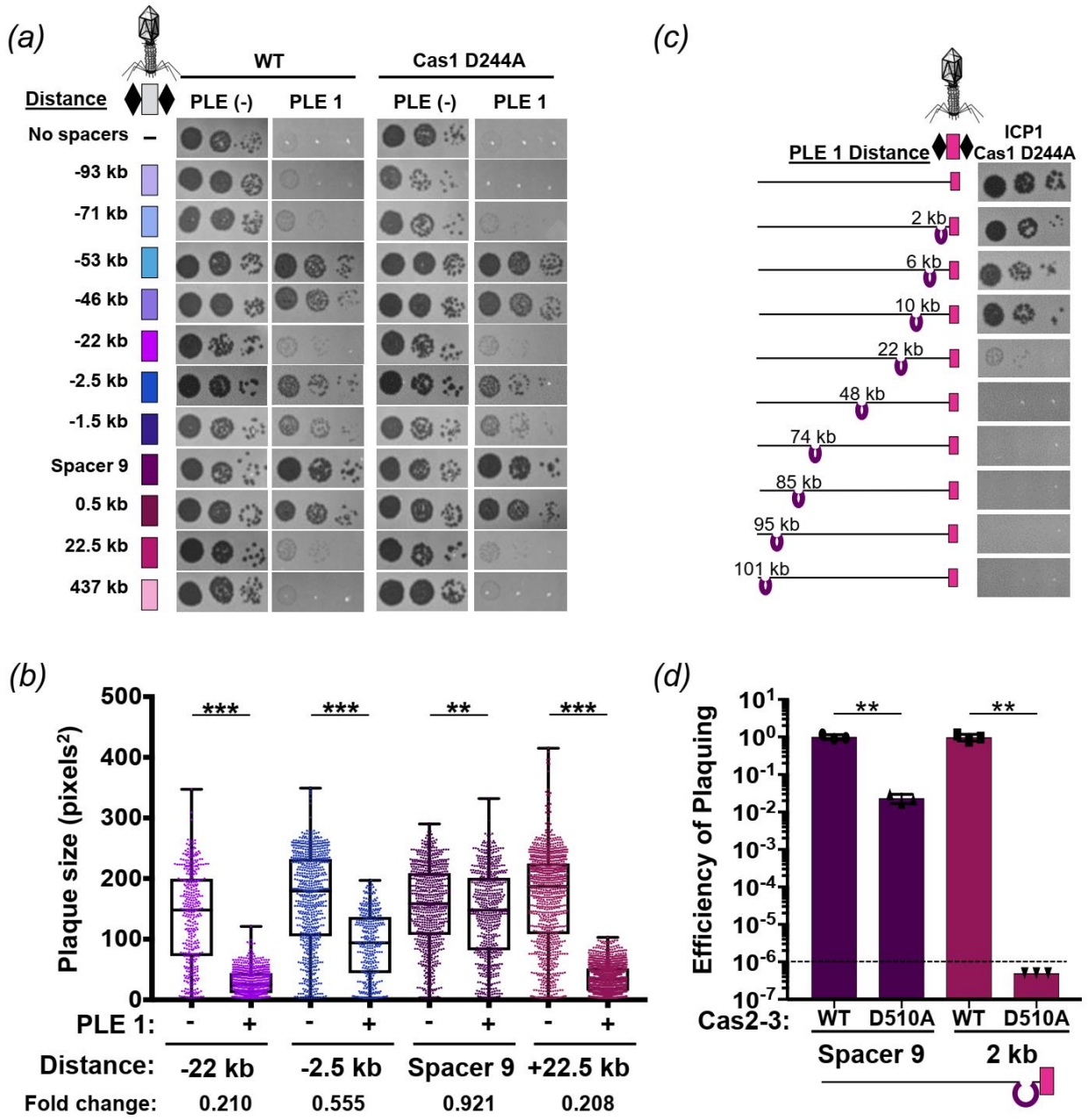


Figure 3.6. The interference potential of spacers directed to the small chromosome is dependent on the proximity to PLE. a, Tenfold dilutions of ICP1 engineered with a spacer that targets the small chromosome or PLE 1 +/- Cas1 spotted on lawns of *V. cholerae* showing the ability of different phage strains to form plaques. Spacer 9 is the same as spacer 9 in ICP1_2011_A. The distance of the chromosomal protospacer from the PLE 1 integration site is indicated. **b,** Plaque size of ICP1 variants plated with *V. cholerae*. The distance and color scheme corresponds to the spacers tested as in **a**. The fold change in average size of a plaque on a PLE (+) host compared to a PLE (-) host is indicated at the bottom. Significance was determined by Mann-Whitney U Test, **p<0.005, ***p<0.0001. **c,** Tenfold dilutions of ICP1 engineered with a chromosomal protospacer spotted on lawns of *V. cholerae* harboring PLE in different locations in the chromosome. **d,** Efficiency of plaquing of phage engineered to contain a spacer that is internal to PLE 1 (spacer 9) or targeting the small chromosome 2kb away from PLE 1 (same as in **c**, cartoon below graph) with the WT or helicase dead (D510A) Cas2-3 allele. The dashed line indicates the limit of detection. Significance was determined by T Test, **p<0.005.

Conversely, ICP1 that target a protospacer only 0.5, 1.5 or 2.5 kb from PLE were able to efficiently overcome PLE and form plaques. Phage targeting protospacers at intermediate distances away from PLE (>20 kb) demonstrated weak plaque formation on PLE 1. Surprisingly, we observed that ICP1 with some spacers targeting relatively far from PLE (53 and 46kb away) were still able to form robust plaques on PLE 1 (Fig. 3.6a). While all of the spacers selected for this assay had one perfect protospacer match in the chromosome and have a GA PAM, we identified >100 promiscuous putative target sites for these spacers which would bring the chromosomal target much closer to PLE 1 (Supplementary Table 3.7), which may explain these phage's ability to overcome PLE. To test if spacer acquisition had a role in plaque formation, we engineered the chromosomal targeting phage in a Cas1 deficient background and assayed for plaque formation on the PLE 1 host. Despite being unable to acquire spacers (Supplementary Fig. 3.6), the Cas 1 deficient phage retained the same plaquing phenotype. We quantified the weaker plaque formation observed when ICP1 targets 2.5kb and >20 kb away from PLE 1 by measuring plaque size compared to PLE (-) *V. cholerae* (Fig. 3.6b). As compared to phage with PLE internal and PLE proximal spacers, phage with chromosomal spacers targeting >20 kb away from PLE had significantly limited plaque size; however, even phage with a chromosomal spacer that is proximal to PLE has a ~50% smaller plaque size when compared to plaques on a PLE (-) host. These results indicate that some PLE-mediated anti-phage activity is retained when CRISPR-Cas is directed at increasing distances from PLE in the small chromosome, but direct targeting of PLE is still required for maximizing phage fitness.

To control for differences in spacer sequences, we also varied the location of the PLE and tested the ability of ICP1 with a single spacer targeting the small chromosome to interfere with PLE 1. Following ICP1-mediated transduction, PLE 1 integrates into a *V. cholerae* repeat (VCR) in the new host [22]. We collected a pool of PLE 1 transductants where PLE was integrated at varying distances from the chromosomal protospacer and challenged these strains with ICP1. As a control, we determined that all of the tested PLE 1 *V. cholerae* hosts were susceptible to ICP1 CRISPR-Cas interference when ICP1 possessed a PLE internal spacer (Supplementary Fig. 3.7a). Consistent with our earlier finding, PLE integrated at an increasing distance away from the protospacer was less susceptible to ICP1-encoded CRISPR interference (Fig. 3.6c).

As Cas2-3 has been demonstrated to translocate *in vitro,* we next wanted to see if the indirect inhibition of PLE by spacers that target the small chromosome was due to Cas2-3 processivity. We constructed a Cas2-3 helicase dead variant by mutating the conserved DExx helicase motif II (D510A) [96] (Supplementary Fig. 3.7b) and tested the ability of ICP1 to form plaques on PLE 1 *V. cholerae*. ICP1 Cas2-3 D510A engineered with a spacer that targets internal to PLE was still able to form plaques on a PLE (+) host, although both the efficiency of plaquing and plaque size were negatively impacted by the helicase mutation (Fig. 3.6d, Supplementary Fig. 3.7c). Conversely, when ICP1 targets the *V. cholerae* small chromosome 2 kb away from PLE, the helicase activity of ICP1 Cas2-3 was absolutely essential for PLE interference and plaque formation (Fig. 3.6d, Supplementary Fig. 3.7c). These findings are the first direct demonstration of functional Cas2-3 processivity *in vivo* and support our model of indirect targeting (Fig. 3.7).

Discussion

Our results reveal that the latest front in the ongoing arms race between contemporary isolates of epidemic *V. cholerae* and its predator ICP1 necessitate the persistence of the ICP1-encoded Type I-F CRISPR-Cas system to counter PLE-mediated anti-phage activity (Fig. 3.1). By using a high-throughput spacer acquisition assay, we gained insight into the full range of spacers that can combat PLE. Interestingly, our experimental findings on acquisition and interference do not reflect the rather limited diversity of spacers that ICP1 maintains against PLE in nature. These results highlight that not all spacers are equally proficient for interference, and that coupled analysis of these competing mobile genetic elements from nature reveals the evolutionary benefits of a particular complement of spacers more so than laboratory-based studies. Despite a lack of clear evidence indicating where the ICP1-encoded CRISPR-Cas system originated, it serves as a tractable model through which we can examine the biology of an endogenous Type I-F CRISPR-Cas system against its cognate foe.

Co-culture studies competing phage against CRISPR-Cas proficient bacterial hosts demonstrated that mutational escape by phage is limited by bacterial populations that have heterogenous CRISPR arrays [104]. Here, we see that PLE 1 is highly conserved over time, even when co-circulating with CRISPR proficient ICP1. In light of previous suggestions, the diversity of CRISPR arrays in ICP1 populations may limit the success of PLE escape mutants. Surprisingly, however, we see very little diversity in the spacer composition of ICP1 CRISPR arrays with the same minimal spacers being conserved in phage circulating for over eight years (Fig. 3.1b). Likewise, CRISPR-proficient ICP1 isolated from nature always encoded more than one spacer against PLE, which would be expected to limit CRISPR escape mutations. It may be that there is limited room for genetic drift in the PLE genome, permitting ICP1 to streamline its CRISPR array, keeping only the most efficient spacers while also maintaining an advantageous genome size.

Akin to studies of bacterial Type I-F CRISPR-Cas mediated interference with plasmid transformation and conjugation [105], we similarly see that the spacer sequence and quantity of spacers in the array have a role in ICP1's ability to abolish PLE spread (Fig. 3.2). This may be due to differences in crRNA abundance or stability, or sequence dependent subtleties that dictate interference potential, as has been proposed previously [106]. Despite spacer 9's improved interference with PLE mobilization compared to spacer 8, we still observed a slight defect in plaque size when comparing engineered phage with only spacer 9 relative to a PLE (-) host (Fig. 3.6b), suggesting that even this improved spacer alone is not sufficient to fully overcome PLE-mediated anti-phage activity. By encoding a seemingly redundant set of spacers targeting PLE, ICP1 increases its ability to overcome PLE and limit PLE spread in the environment. Additionally, multi-site targeting of *V. cholerae* PLE by ICP1 CRISPR-Cas may contribute to the modular evolution observed between PLE variants and dictate which PLEs are circulating within and between epidemics [22].

As expected, the majority of spacers acquired in our high-throughput acquisition assay directly target PLE (Fig. 3.3a). Analysis of natural ICP1 isolates recovered from cholera patient stool samples shows that the phage-encoded CRISPR-Cas system recognizes a GA PAM, (Supplementary Fig. 3.4) which, although atypical for Type I-F

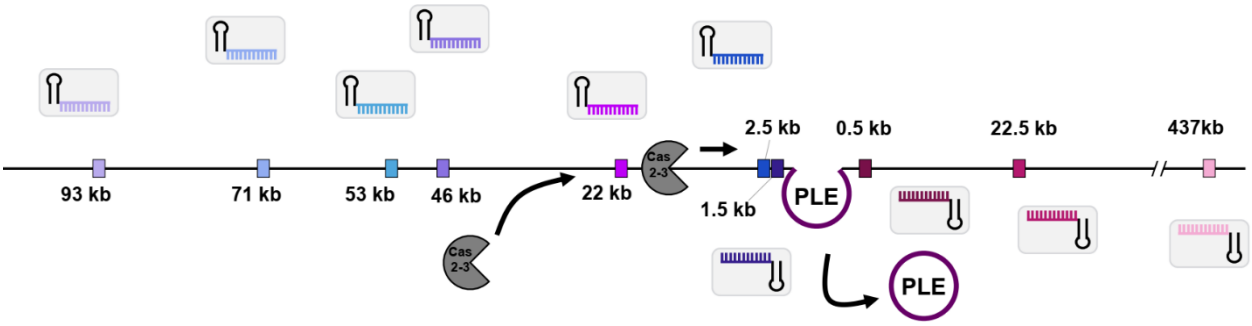


Figure 3.7. Model of race between ICP1 Cas2-3 processive degradation of the *V. cholerae* chromosome and ICP1-mediated PLE excision. Csy complexes (grey boxes) with crRNAs (colored) search for a complementary protospacer (colored rectangles, experimentally assessed in Fig. 3.6a). Cas2-3 (dark grey) is recruited to the protospacer and processively degrades the DNA towards PLE (purple). ICP1 is able to form plaques when Cas2-3 degrades PLE before PLE excises from the chromosome, which occurs within 5 minutes of ICP1 infection.

systems [107], has been confirmed through single mutations to a C in both positions [24]. Notably, we found that ICP1 was able to incorporate spacers that targeted non-canonical PAMs (Fig. 3.3d) and that these spacers can suffice for PLE interference (Fig. 3.4b). In comparison to another high throughput spacer acquisition assay in a Type I-F system, which found >90% of all protospacers flanked by the canonical PAM [101], it appears that the phage-encoded system is less discriminating with only 70% of protospacers flanked by the expected PAM. However, targeting a protospacer with a non-canonical PAM reduced the ability of ICP1 to form plaques compared to the canonical PAM (Fig. 3.4b). As such, in nature ICP1 targeting a protospacer with a non-canonical PAM would not be able to completely interfere with PLE and thus would be selected against. This hypothesis is additionally supported by the observation that very few non-canonical PAM protospacers were associated with indirect targets in the small chromosome. As these chromosomal spacers are themselves less proficient for interference (Fig. 3.6a and 6b), the added disadvantage of targeting a protospacer with a non-canonical PAM likely tips the balance in favor of PLE, likely explaining the lower abundance of these spacers in our selection experiments.

Despite the presence of spacers that target the *V. cholerae* large chromosome in the high-throughput spacer acquisition assay (Fig. 3.3c), we show that targeting this chromosome is dispensable for CRISPR interference of PLE and is likely an artifact of a duplication event of the mu-like region in the strain used in our assays (Fig. 3.5). Interestingly, two of the natural ICP1 isolates contain a spacer that targets a gene on the *V. cholerae* large chromosome (Fig. 3.1b). We speculate that this spacer was acquired from a *V. cholerae* strain possessing a duplication or rearrangement that is not represented in currently sequenced isolates, in which the protospacer was in the small chromosome proximal to PLE, allowing the phage to overcome PLE activity. However, this spacer does not seem to be maintained in the phage population, likely due to diminished PLE interference relative to PLE-direct spacers as we experimentally observed.

CRISPR targeting of the *V. cholerae* small chromosome can overcome PLE, but our results suggest a model in which there is a limit to the distance over which processive Cas2-3 degradation can occur to reach the PLE prior to excision (Fig. 3.7), an action which occurs within five minutes of ICP1 infection that is directed by an early-

expressed ICP1 protein [29]. The limit of processivity appears to be around a distance of 23 kb (Fig. 3.6a and 6c), at which point either Cas2-3 is unable to continue to process along the *V. cholerae* chromosome or PLE excises before interference occurs. *In vitro* studies of Cas3 from Type I-E systems have demonstrated Cas3 translocation velocities of 89 to 300 bases per second and average processivities between 12 to 19 kb [108,109], however, the functional role and limitations of processivity *in vivo* are not known. Our results are the first to demonstrate that Cas2-3 is processive *in vivo,* with over 22 kb from a distal chromosomal protospacer over which the ICP1 CRISPR-Cas can maintain activity to overcome PLE when Cas2-3 has the ability to translocate along DNA (Fig. 3.6d). As this event must occur within five minutes of ICP1 initiating infection, the estimated processivity of ICP1 Cas2-3 is within the range of what has been reported for Type I-E Cas3, which is especially remarkable given the complexity of the crowded intracellular environment compared to simplified *in vitro* systems.

In comparison to other Cas nucleases like Cas9, which introduces a single double-stranded break [110,111], Cas2-3 degrades DNA as it translocates away from the protospacer [97], making it more likely to destroy and thus interfere with its target. In fact, we see that the helicase dead Cas2-3 is less able to overcome PLE even when directly targeting the anti-phage island, suggesting that the processive degradation of PLE contributes to interference (Fig. 3.6d). Similarly, this predicted advantage may account for the increased prevalence of Type I systems for phage defense [112]. In the context of the battle between ICP1 and PLE, this processivity permits interference even with an indirect CRISPR target and has important implications for harnessing CRISPR-Cas in biotechnology and medicine. Since the characterization of the ICP1-encoded CRISPR-Cas system, phage engineered with CRISPR-Cas systems to target virulent, antibiotic resistant bacteria have been assayed for therapeutic applications [113,114], showing the value of innovating from natural systems to overcome disparate biological problems.

Methods.

Strains, growth conditions and genomic analysis.

Phage, bacterial strains and plasmids used in this study are listed in Supplementary Tables 3.1-3.3. Bacteria were routinely grown at 37℃ on lysogeny broth (LB) agar or in LB broth with aeration. Media was supplemented with ampicillin (50 μg/ml), kanamycin (75 μg/ml), spectinomycin (100 μg/ml), and/or streptomycin (100 μg/ml) when appropriate. Phage susceptibility was determined by standard soft agar overlays as described [53] and phage plaque spot plates were performed as described previously [29]. Images are representative of at least two independent assays. Cholera stool samples collected and stored at the ICDDR,B between 2015-2017 were probed for the presence of phage by standard soft agar overlays, and *V. cholerae* isolates were recovered by plating on Thiosulfate Citrate Bile Salts Sucrose selective media (Difco). ICP1 specific primers [19,24] and PLE specific primers (Supplementary Table 3.4) were used for preliminary screening of isolates from stool samples. The presence of CRISPR-Cas in ICP1 and PLE in *V. cholerae* was validated by whole genome sequencing. Genomic libraries were generated using NEBNext Ultra II DNA Library preparation kit for Illumina (New England Biolabs), according to the manufacturer's recommended protocols. Paired-end sequencing (2 × 150 bp) was performed on an Illumina HiSeq4000 (University of California, Berkeley QB3 Core Facility). Sequencing assembly/mapping

and detection of CRISPR was performed as described [21]. The genome of the *V. cholerae* clinical isolate KS393 was sequenced on Illumina HiSeq4000, PacBio Sequel and Oxford Nanopore MinION sequencers (University of California, Berkeley QB3 Core Facility). Assembly of KS393 sequences was performed using the canu assembler v1.6 [115] to combine the PacBio and Oxford Nanopore reads into genomic scaffolds for the large and small chromosomes using default settings and an expected genome size of 4033460bp. This generated two scaffolds of the expected sizes for each chromosome which were then polished with the Illumina paired-end sequences using Pilon v1.22 [116] with the "fix all" command to generate a high-quality genomic assembly in a fasta format of both chromosomes (Supplementary File 1). The sequencing data for strain KS393 have been deposited in the Sequence Read Archive (SRA) database under accession codes SRR7826356, SRR7826357 and SRR7826358.

V. cholerae mutants were constructed by natural transformation as described [50]. Mutations in ICP1 were generated using CRISPR-Cas mediated genome engineering with the *V. cholerae* classical biotype Type I-E system as described [53] (Supplementary Tables 3.3 and 3.4). Engineered phage +/- Cas1 D244A with spacer 9 were validated by plaquing on a permissive PLE 1 host and determining the frequency of phage with a newly acquired spacer by calculating the efficiency of plaquing on the permissive PLE 1 host to a PLE 1 host with the protospacer deleted (PLE $1^{\Delta PS9}$). The limit of detection is met when the phage is unable to form a plaque on the restrictive host at the highest concentration while still being able to productively infect a permissive host, with at least 6 orders of magnitude tested. Examination of PLE replication and transduction during phage infection was described as reported previously [22].

High throughput spacer acquisition, data processing and analyses

Three independent experiments were performed as follows: A 50 mL culture of PLE 1 *V. cholerae* was grown to OD_{600} = 0.3 and infected at an MOI of 1 with ICP1_2011_A ΔS9, which harbors spacer 8 at the leading edge of the CRISPR 1 array that targets PLE 1 and allows for phage replication [24]. Infected cells were incubated for 90 minutes at 37℃ with aeration, at which point lysis was observed. The lysate was treated with chloroform and centrifuged to remove bacterial debris. Phage were precipitated with 10% (w/v) polyethylene glycol (PEG) 8000 at 4°C overnight. Phage pellets were collected by centrifugation at 4°C and the passaging was repeated as above. After three passages, the resulting pools were plated on a host possessing silent mutations in protospacer 8 (PLE 1^{PS8*}) that inhibits spacer 8-mediated CRISPR interference, which enabled the selection of phage with expanded arrays that allow plaque formation. Phage DNA libraries were generated by homopolymer tail-mediated PCR (HTM-PCR) as previously described [117]. As ICP1_2011_A possesses only a single functional CRISPR array (Fig. 3.1b), the expanded phage CRISPR 1 array was amplified from genomic DNA libraries by PCR using custom barcoded primers (Supplementary Table 3.4) to sequence the leader proximal spacer. 50bp single-end sequencing was performed on an Illumina HiSeq 2500 (Tufts University Core Facility) using a custom sequencing primer. The resulting reads as fastq files were mapped to the large and small chromosome of *V. cholerae* strain KS393 using Bowtie v1.2.2 [118] with a seed_length of 31 and allowing for 0 max_total_mismatches which ensured that spacer to protospacer matches were 100% identical. These mappings were performed in two parallel ways: first, to obtain all

possible spacer mapping locations regardless of the number of identical protospacer targets (i.e. translucent spacers in Fig. 3.3b) and second, restricting max_alignments to 1 which only mapped spacers with exactly one unique mapping location across both chromosomes. With a custom Python script (https://git.io/fNVqZ) we extracted the PAM sequences and GG PAM slippage locations from the restricted unique mappings. We also used this script to generate spacer mapping location graphs for both set of mappings using Biopython's GenomeDiagram module [119]. The amplicon sequencing data have been deposited in the SRA database under accession codes SRR7827053, SRR7827054, SRR7827055.

Acknowledgments
The authors are especially thankful to ICDDR,B hospital and laboratory staff for their support and would like to thank Shirajum Monira, Kazi Zillur Rahman, Fatema-tuz Johura, Marzia Sultana, and Monika Sultana in particular. The authors would also like to thank members of the Seed lab for critical feedback and thoughtful discussion regarding this manuscript.

Chapter 5

The arms race between host and virus goes on

Through extensive analysis, we are now able to update our model of ICP1 infection of a PLE (+) *V. cholerae* cell. Upon infection, ICP1 produces both PexA and HelA, the former of which physically interacts with the PLE-encoded, constitutively expressed Int. The interaction of Int with PexA directs Int to catalyze the PLE excision and circularization reaction within 5 minutes of infection, thus allowing PLE to escape from the host chromosome. PLE then hijacks HelA to facilitate PLE replication to high copy, allowing for increased transcription of PLE-encoded genes and interference with ICP1 replication. Mobilization of PLE out of the chromosome thus occurs before ICP1 expresses its phage-encoded nucleases that degrade the *V. cholerae* chromosome, protecting PLE from ICP1 host takeover. Conversely, when ICP1 has a functional CRISPR-Cas system, PLE is susceptible to degradation by the processive Cas2-3 helicase/nuclease both when integrated in the chromosome and indirectly targeted, as well as when PLE is excised and replicating and directly targeted by CRISPR-Cas. Increased efficiency at interfering with PLE by directly targeting the PLE itself has led to selection of ICP1 CRISPR arrays with multiple spacers targeting PLE, diminishing the ability of PLE to replicate and transduce. Altogether, we demonstrate overarching interactions and coevolution that have implications on the interactions of hosts and viruses that can be applied to all levels of the tree of life.

The characterization of PexA as the RDF for the PLE-encoded Int is the first example of an integrase and RDF being encoded in separate genomes. Often, bioinformatic analysis of genomic islands and lysogens are characterized by the presence of an integrase; however, the absence of an RDF leads to the prediction that the island or phage is degenerate and unable to be mobilized. The revelation that genomic islands, such as PLE, can hijack phage-encoded proteins to direct mobilization shifts the paradigm of degenerate islands, and, instead, provides more possibilities of directed mobilization from infecting phages. As PexA is specific to ICP1, it is the first piece of molecular evidence that PLE has evolved to respond to and defend against ICP1 infection. It is remarkable that PexA has no predicted homologs in any sequenced phages, including the marine phages that have SF1B type helicases. The fact that it is highly conserved in all ICP1 isolates suggests that it plays a critical role in the ICP1 lifecycle, although the function may be obscured by the lab conditions under which it was examined. To discern the role that PexA has in the ICP1 lifecycle, biochemical approaches will be needed to determine what, if any, *V. cholerae* or ICP1 proteins it interacts with, as well as analysis in varied media conditions that mimic either environmental reservoirs or mammalian gut environments. As more phages are sequenced and added to genome databases, it additionally will be interesting to see if any homologs of *pexA* are discovered to hint at the evolution and function of this highly specific ICP1 gene product.

Of all the PLEs that have been characterized, PLE 2 is the only one that does not excise in response to *pexA* expression. As it does circularize during infection, it stands to reason that PLE 2 has evolved to recognize another ICP1-encdoed gene product to direct its excision. In addition to screening ICP1 mutants for defects in PLE 2 excision, it would also be interesting to examine the evolutionary trajectory that led PLE 2 to acquire a more divergent integrase.

PLE mobilization is critical for PLE escape from ICP1 host takeover, as ICP1 encoded nucleases are able to degrade PLE when it is stuck in the *V. cholerae*

chromosome. In addition to excision directed by PexA, we also show that PLE makes use of ICP1-encoded HelA, an accessory helicase, to drive PLE replication, mediating escape from ICP1 degradation of the *V. cholerae* chromosome. PLE's use of ICP1 HelA for replication is unique compared to PICIs that autonomously replicate using host-encoded machinery. Additionally, the requirement for ICP1 infection in addition to *in trans* expression of *helA* suggests that PLE needs other components of ICP1's replication machinery to increase copy number. This requirement makes PLEs much more reminiscent of the satellite viruses of plants that make use of helper RNA virus-encoded RNA-dependent RNA polymerase to replicate during helper virus infection[120], emphasizing the similarities between viral-host systems across different kingdoms of life.

Another similarity characteristic of viral infection is the necessity to overcome and subvert host processes to favor progression through the viral program. The ability of ICP1 to overcome PLE when PLE is unable to excise from the chromosome or replicate is most likely due to phage-encoded nucleases that recycle the host chromosome during infection. ICP1 is predicted to encode over 10 putative endonucleases that could be responsible for chromosomal degradation during infection either independently or in tandem with a phage-encoded exonuclease. Ideally, identification of the endonuclease(s) responsible for chromosomal degradation would confirm the role of chromosomal recycling in allowing ICP1 to overcome replication and excision deficient PLE.

More common methods for viruses to overcome their hosts is through modulation of transcription or transcript levels to favor transcription from the infecting virus[79,80]. A common practice for phages that do not encode RNA polymerase is to encode alternative sigma factors that remodel the interaction between the DNA and RNA polymerase to force transcription of phage genes, referred to as sigma appropriation[63]. ICP1 is not predicted to encode an RNA polymerase, although it remains to be seen what kind of effect ICP1 has on subverting *V. cholerae* transcription machinery. In order for PLE to be effective against ICP1, it must be transcriptionally active throughout the course of ICP1 infection. This requirement sets up another interesting line of inquiry into host takeover evasion, and more work is needed to identify if PLE encodes transcription specific genes to favor PLE transcription over ICP1, or if makes use of ICP1-like promoters to potentially make use of ICP1-mediated sigma appropriation.

While *helA* is necessary for PLE replication and has a role in ICP1 fitness, the exact mechanism by which *helA* is used by either entity remains to be further elucidated. The universal requirement for *helA* in replication of all PLEs suggests that HelA is being recruited to either a conserved PLE DNA sequence or by a conserved PLE-encoded protein, such as RepA, which has a highly conserved C-terminus and similar phenotypes to ICP1 *ΔhelA* when knocked out. The fact that *helA* is not necessary for ICP1 replication but still contributes to phage fitness suggests that it has a role outside of the replisome, potentially in maintaining genome integrity. Such a role would not have an obvious phenotype in a single round of replication, but could lead to accumulation of non-viable virions that would be inefficient at infecting new hosts, perhaps explaining the observed smaller plaque phenotype in ICP1 *ΔhelA,* and the high level of conservation of an SF1B type helicase in sequenced ICP1 isolates.

Despite the difficulty that ICP1 has in overcoming PLE *in vitro*, ICP1 is quite readily co-isolated from cholera patient stool samples along with PLE (+) *V. cholerae* due to its ability to target and degrade PLE via the phage-encoded CRISPR-Cas system. In response to an increase in the prevalence of PLE (+) *V. cholerae* being shed in epidemics, all ICP1 isolates recovered from recent epidemics in Dhaka, Bangladesh encode a CRISPR-Cas system with most, if not all, of the spacers targeting PLE. Despite the ability of the ICP1 CRISPR-Cas system to indirectly interfere with PLE by targeting the *V. cholerae* chromosome and Cas2-3 processively degrading the DNA until PLE is cut, it is less favorable than directly targeting PLE because PLE is, in a sense, a moving target that excises from the chromosome and replicates.

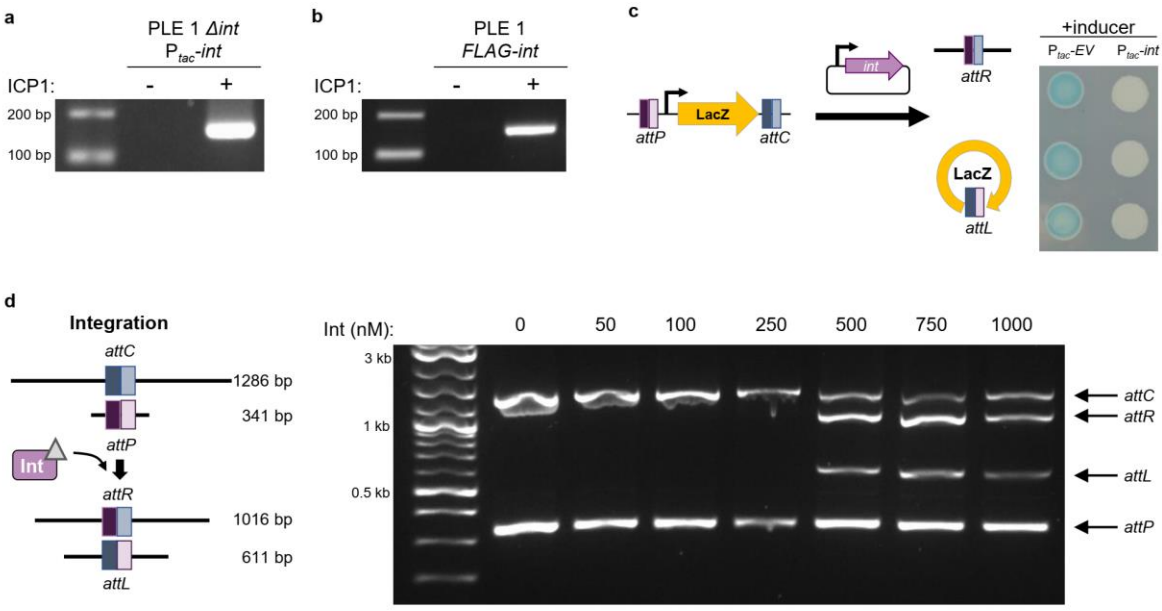
Remarkably, while a single ICP1 encoded CRISPR spacer is sufficient to allow for plaque formation on a PLE (+) host, a single spacer is not sufficient to inhibit PLE replication and PLE transduction—in fact, it is not until ICP1 encodes at least three unique spacers against PLE that it is completely able to abolish PLE replication. This data is counterintuitive to the way in which CRISPR-Cas systems are thought to be a binary all-or-nothing defense system, with single spacers providing population-wide ability to overcome PLE but in single rounds of infection not completely shutting down the anti-phage program. Notably, two different ICP1-encoded spacers that were naturally acquired showed vastly different abilities to interfere with PLE replication and transduction, although the reasoning behind the different levels of inhibition is unknown. It would be interesting to see if the properties of the spacer, such as the GC content, influences the ability to interact with PLE protospacers or contributes to spacer stability, thus altering the ability to interference with PLE.

As the majority of spacers that have been mapped from clinical isolates of ICP1 target PLE ORFs, it also raises the question of the role that CRISPR-Cas targeting has in the modular evolution of PLE. No mutations were detected in PLEs that were able to transduce following infection with a CRISPR-Cas proficient phage. Additionally, there is no evidence for accumulation of mutations in PLE over time. Instead, PLEs recombine different sections with, presumably, other PLEs and mobile genetic elements in the *V. cholerae* chromosome to evolve in a modular fashion over time. Thus, in order to coexist with CRISPR-Cas (+) ICP1, perhaps PLE is able to swap out regions that are targeted by CRISPR in order to thrive during targeting. This new mosaic PLE would then be able to escape ICP1 CRISPR-Cas targeting and could take over the epidemic population.

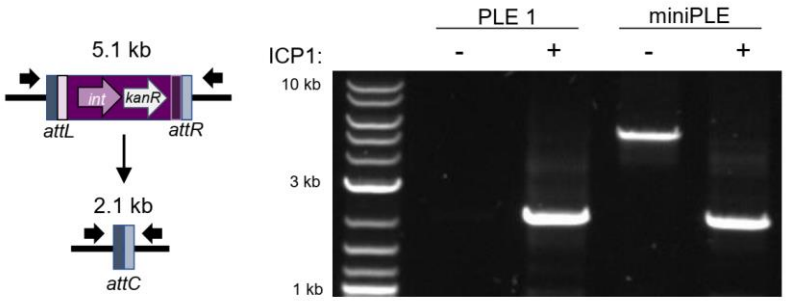
Altogether, the predation of *V. cholerae* by ICP1 has significant implications on the evolution of this important bacterial pathogen. In light of the threat of ICP1 infection, epidemic *V. cholerae* has acquired the anti-phage defense island, PLE, that is mobilized by ICP1 infection thought direct interactions with phage-encoded proteins PexA and HelA. Not only does this mobilization facilitate HGT, but it also allows the anti-phage island to overcome host takeover mechanisms endogenous to ICP1. While CRISPR targeting of PLE by ICP1 does allow phage to productively infect the PLE (+) *V. cholerae*, PLE is not fully inhibited and still is able to mobilize. Indeed, if CRISPR targeting of PLE helps to facilitate recombination with mobile genetic elements within the chromosome, then infection with ICP1 could lead to evolution of the anti-phage phage island into more of a phage-mobilized pathogenicity island, like SaPIs. Such a scenario is especially dire, in light of the proposed use of ICP1 as a therapeutic to treat

epidemic cholera[48]. Thus, the insights gained from the study of the tripartite ICP1-PLE-V. cholerae system not only add to our understanding of the evolution of epidemic *V. cholerae* but also broaden our understanding of the layers of molecular parasitism plaguing microbial interaction in the context of human disease.

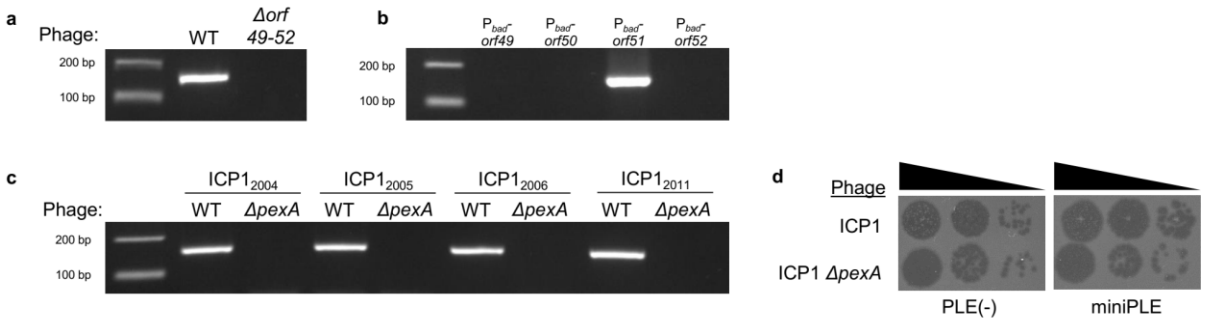
Supplementary Information



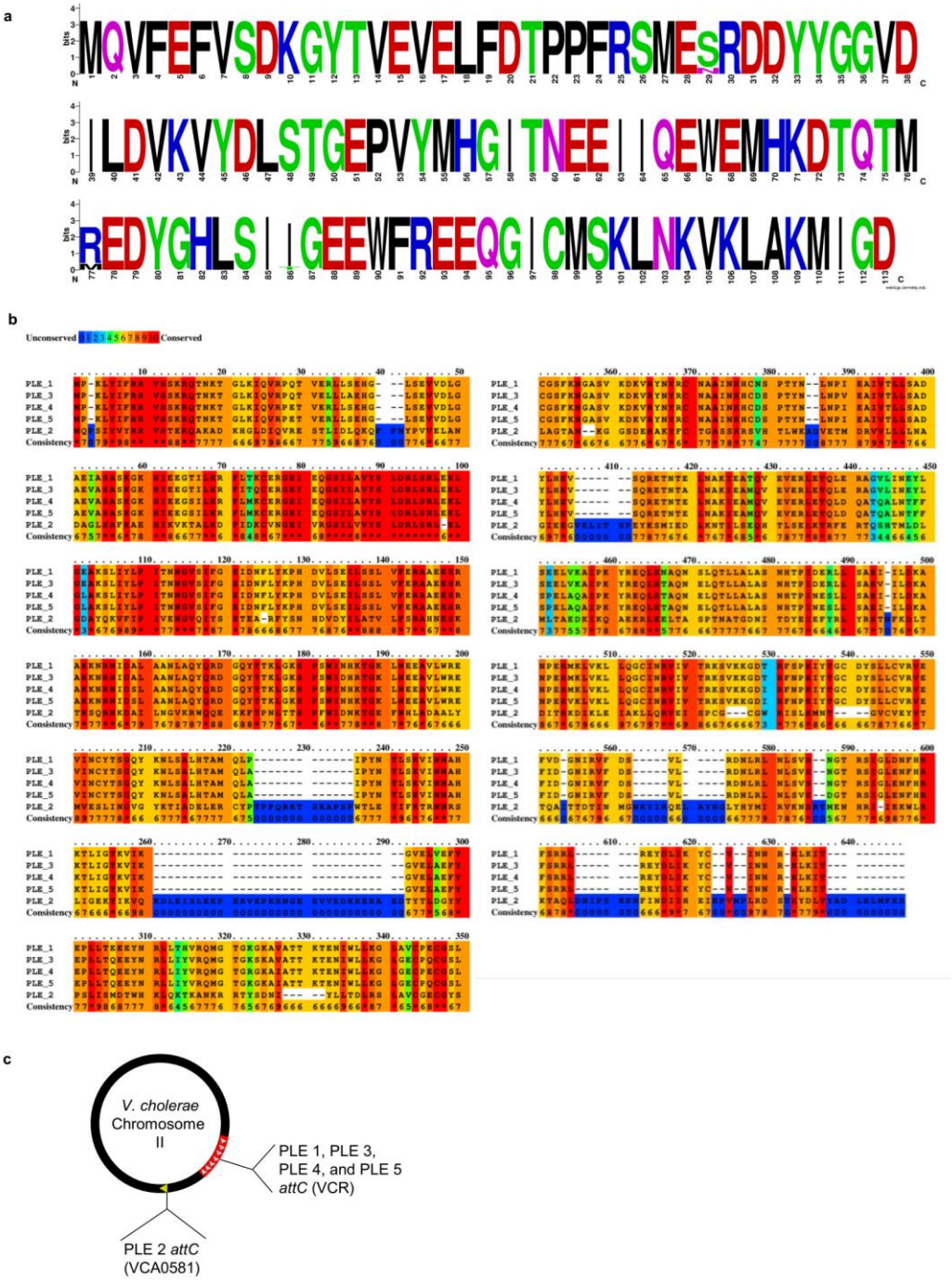
Supplementary Figure 1.1. PLE 1 Int is a functional integrase. a, Ectopic expression of Int is not sufficient to drive PLE circularization in the absence of ICP1 infection. **b,** Endogenously FLAG-tagged PLE Int is still able to drive functional circularization within 5 minutes of ICP1 infection. **c**, Cartoon depicting the *in vivo* integration assay. Constitutively expressed LacZ from *E. coli* was flanked by PLE 1 *attP* and *attC* sites and integrated into the *V. cholerae* chromosome. When ectopically expressed, Int recombines the *att* sites, creating the hybrid *attL* and *attR* sites and excising LacZ from the chromosome. Three independent colonies of each vector type were spotted on indicator plates containing IPTG, theophylline, and X-gal and incubated for 24 hours. **d**, Left, cartoon depicting an *in vitro* integration assay, in which dsDNA containing the PLE 1 *attC* and *attP* sites is incubated with Int to determine if the recombination products *attL* and *attR* are produced. Right, results of the *in vitro* integration assay showing increasing concentrations of Int lead to the production of *attR* and *attL*.



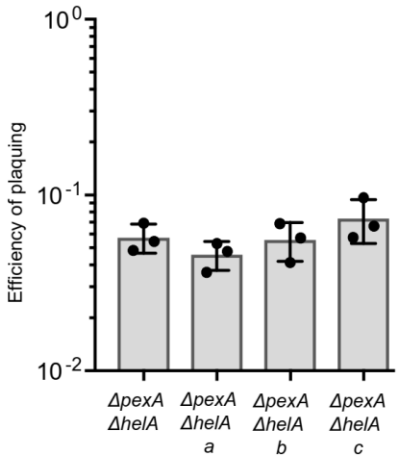
Supplementary Figure 1.2. The miniPLE excises from the chromosome during ICP1 infection. Left, cartoon of primers used to detect PLE and miniPLE excision. Right, PCR to detect excision of PLE 1 or miniPLE during ICP1 infection. PLE 1 is 18kb and cannot be PCR amplified when integrated in the genome.



Supplementary Figure 1.3. ICP1-encoded ORF51 (PexA) contributes to PLE circularization. a, miniPLE circularization plaque PCR. Agar stabs from ICP1$_{2004}$ plaques on miniPLE were boiled and used as template to detect circularization. When ICP1$_{2004}$ has a mutation in *orf49-52*, miniPLE circularization is not detected. **b,** Individual ORFs were cloned into an inducible P$_{bad}$ vector and screened for miniPLE circularization. PLE circularization is detected only when *orf51* is induced. **c,** ICP1 *pexA* is necessary for miniPLE circularization in multiple ICP1 isolates. Different IPC1 isolates with and without clean deletions of *pexA* were tested for miniPLE circularization via plaque PCR. **d,** Tenfold dilutions of ICP1 spotted on various *V. cholerae* lawns shows that ICP1 *ΔpexA* does not have any defects in plaque formation (dark spots, zones of killing) relative to wild-type ICP1.

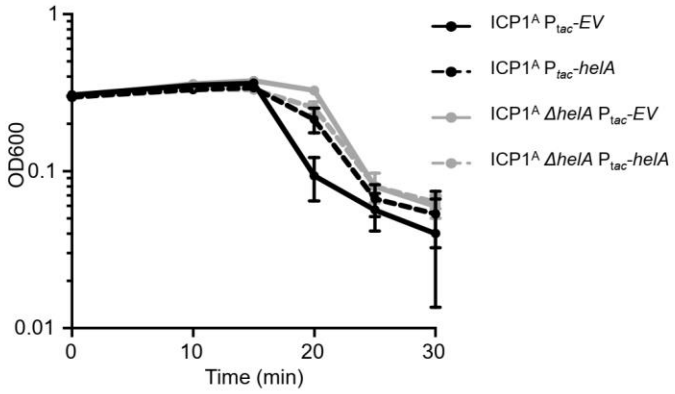


Supplementary Figure 1.4. The conservation of PLE Int and ICP1 PexA between isolates. a, Sequence logo[57] depicting the amino acid sequence of PexA from 17 ICP1 isolates (Supplementary Table 1.5) from between 2001 and 2012. **b**, Praline alignment[58] of amino acid sequence conservation across Int from the characterized PLEs[22]. **c**, Cartoon depicting *attC* sites of characterized PLEs. PLE 1, PLE 3, PLE 4, and PLE 5 all integrate into the *V. cholerae* repeat[22] (VCR, grey triangles), of which there are over 100 sites within the super-integron[100] (red). PLE 2 integrates into a unique site in gene VCA0581[22] (yellow triangle) that shares no sequence similarity with the VCR.

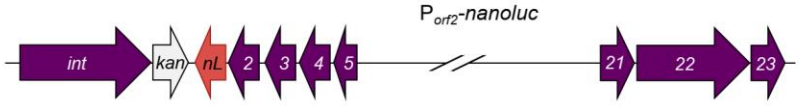


Supplemental Figure 2.1. Three sets of plaques from ICP1 *ΔpexA ΔhelA* on PLE (+) *V. cholerae* were picked and the efficiency of plaquing on PLE (+) relative to PLE (-) *V. cholerae* was retested.

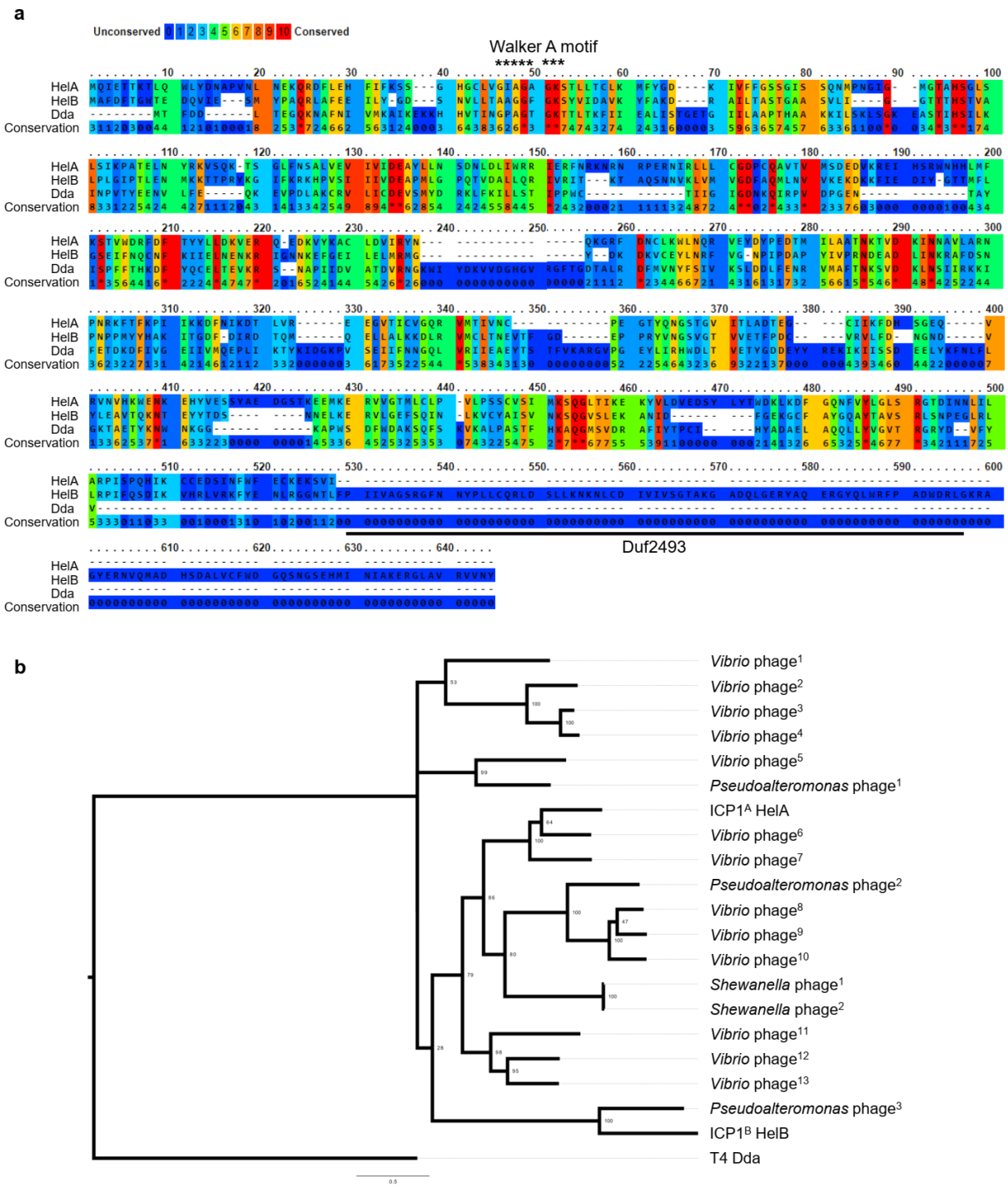
a



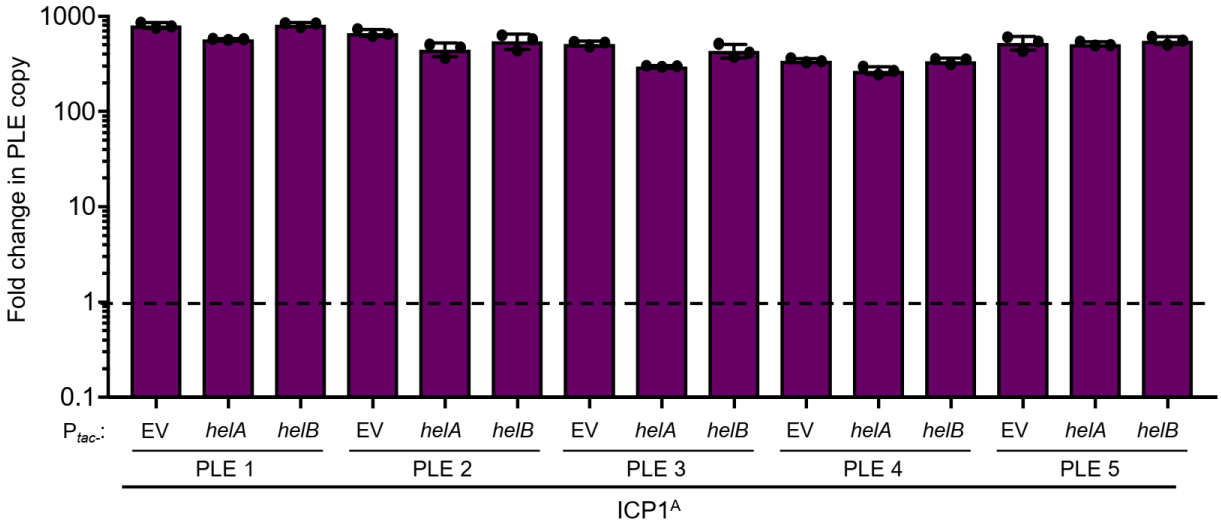
b



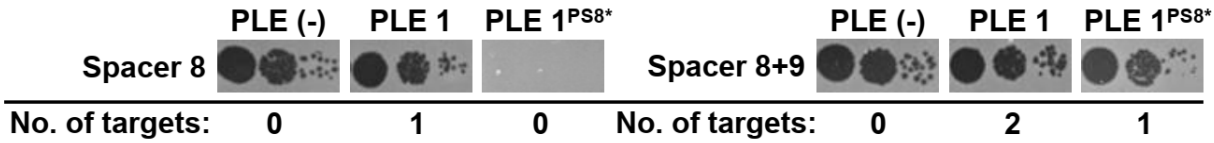
Supplemental Figure 2.2. a, OD_{600} of the listed PLE (+) *V. cholerae* following infection with the listed ICP1. **b,** Cartoon of the PLE-encoded nanoluciferase reporter used, with nanoluciferase encoded downstream of PLE *orf2.*



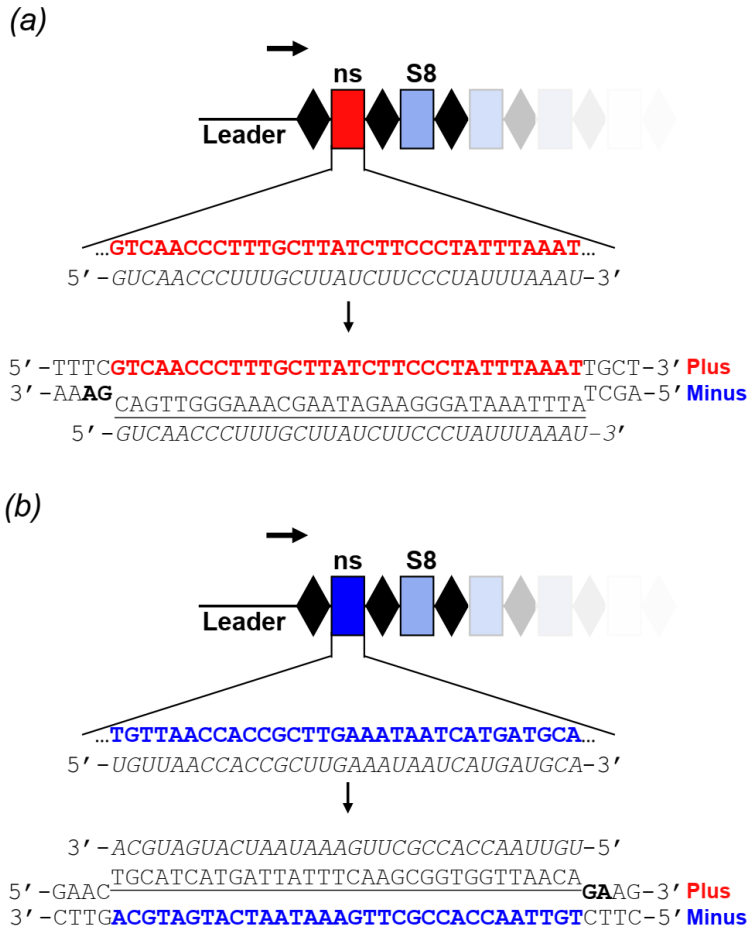
Supplemental Figure 2.3. a, Praline alignment [58] of ICP1^{A} HelA, ICP1^{B} HelB, and T4 Dda. The Walker A motif used in ATP hydrolysis is indicated by asterisks, and the Duf2493 in HelB is marked. **b,** Phylogenetic analysis of SF1B-type helicases, listed in Supplementary Table 2.7.



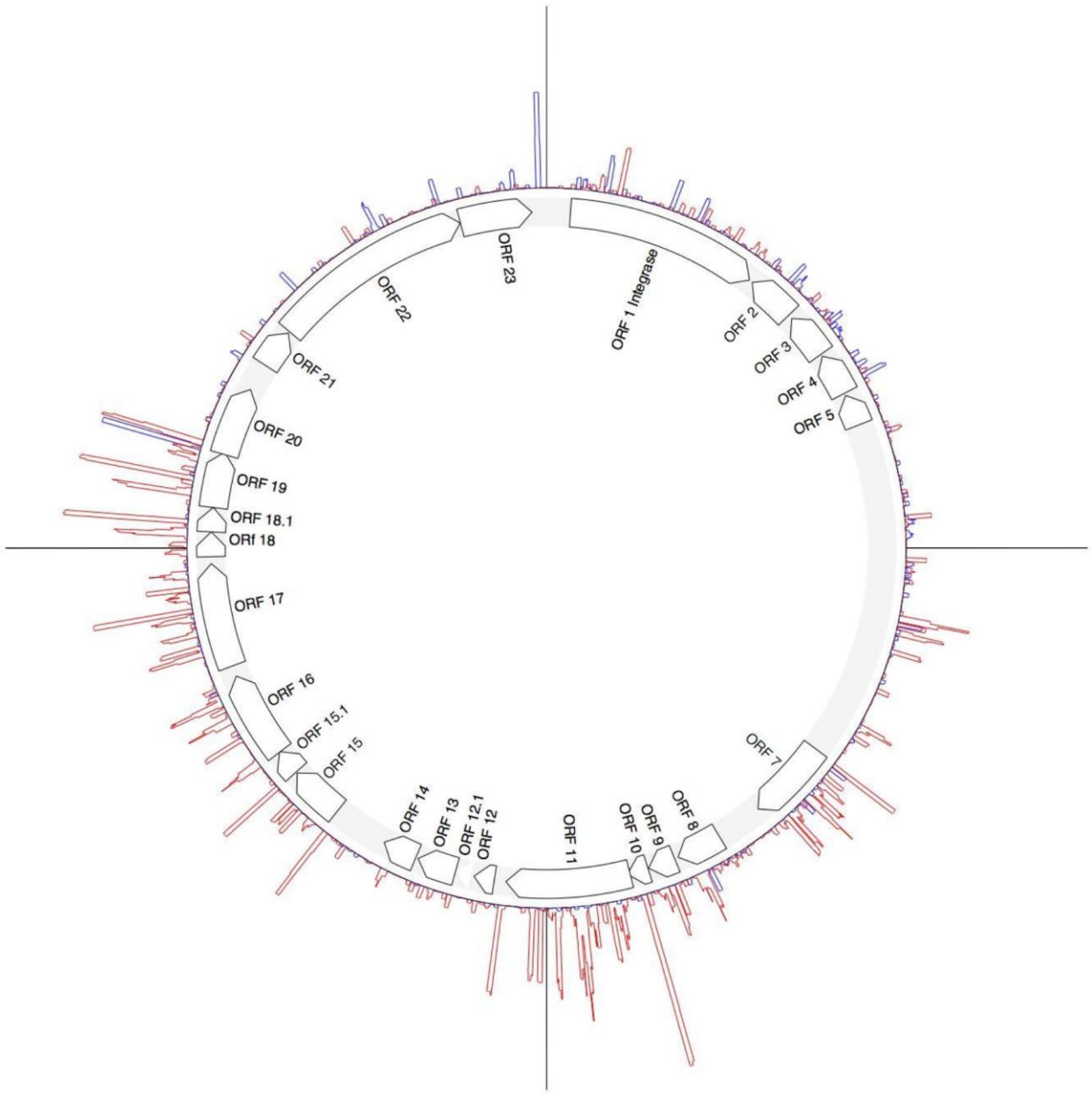
Supplemental Figure 2.4. Replication of the listed PLE in isogenic hosts 20 minutes following infection by ICP1^{A}. Ectopic vectors were induced 20 minutes prior to infection. Dashed line indicates no change in copy.



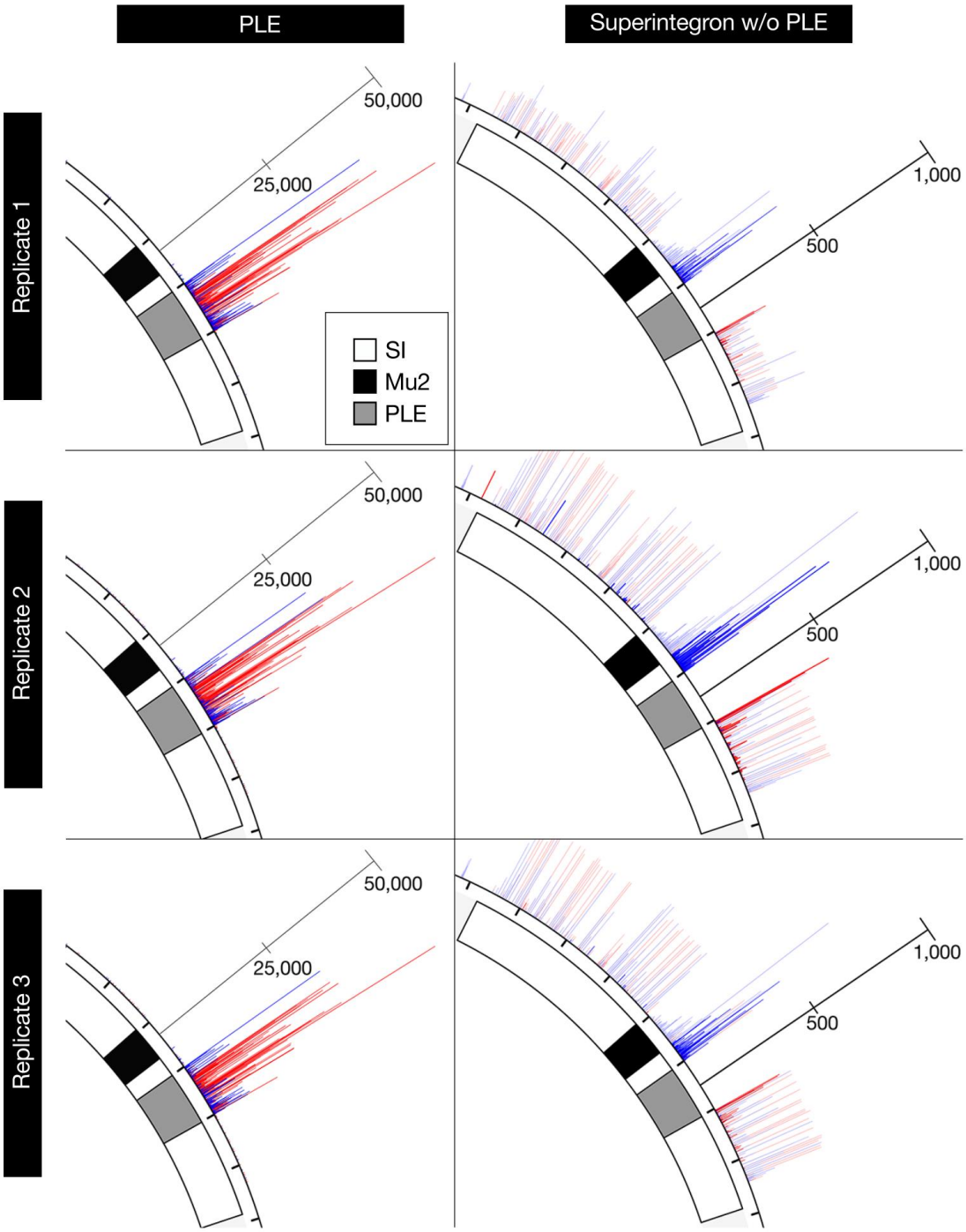
Supplementary Figure 3.1. A single spacer is necessary and sufficient to permit lytic growth of ICP1 on PLE 1 *V. cholerae* as seen by equal plaque formation. Tenfold dilutions of ICP1 with the indicated spacers were spotted on the labeled *V. cholerae* lawns. PS8* indicates silent mutations in protospacer 8 that abolishes CRISPR interference [24]. The number of functional protospacers are listed.



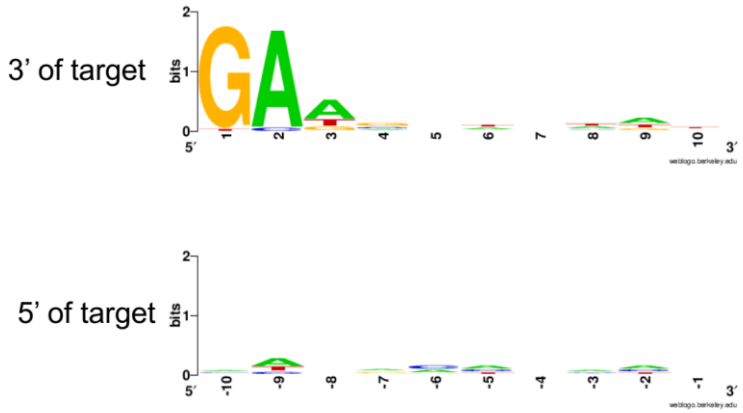
Supplementary Figure 3.2. Schematic of high-throughput spacer acquisition mapping to the *V. cholerae* chromosome(s). The newest spacer from the ICP1 CRISPR array was sequenced using the oligo indicated in the black horizontal arrow, with the 50bp single end illumine read initiating from that arrow. The new spacer either matches to the plus strand *(a)* or the minus strand *(b)*. The spacer DNA sequence is bolded and colored, the crRNA is italicized, the protospacer is underlined and the PAM is bolded in black.



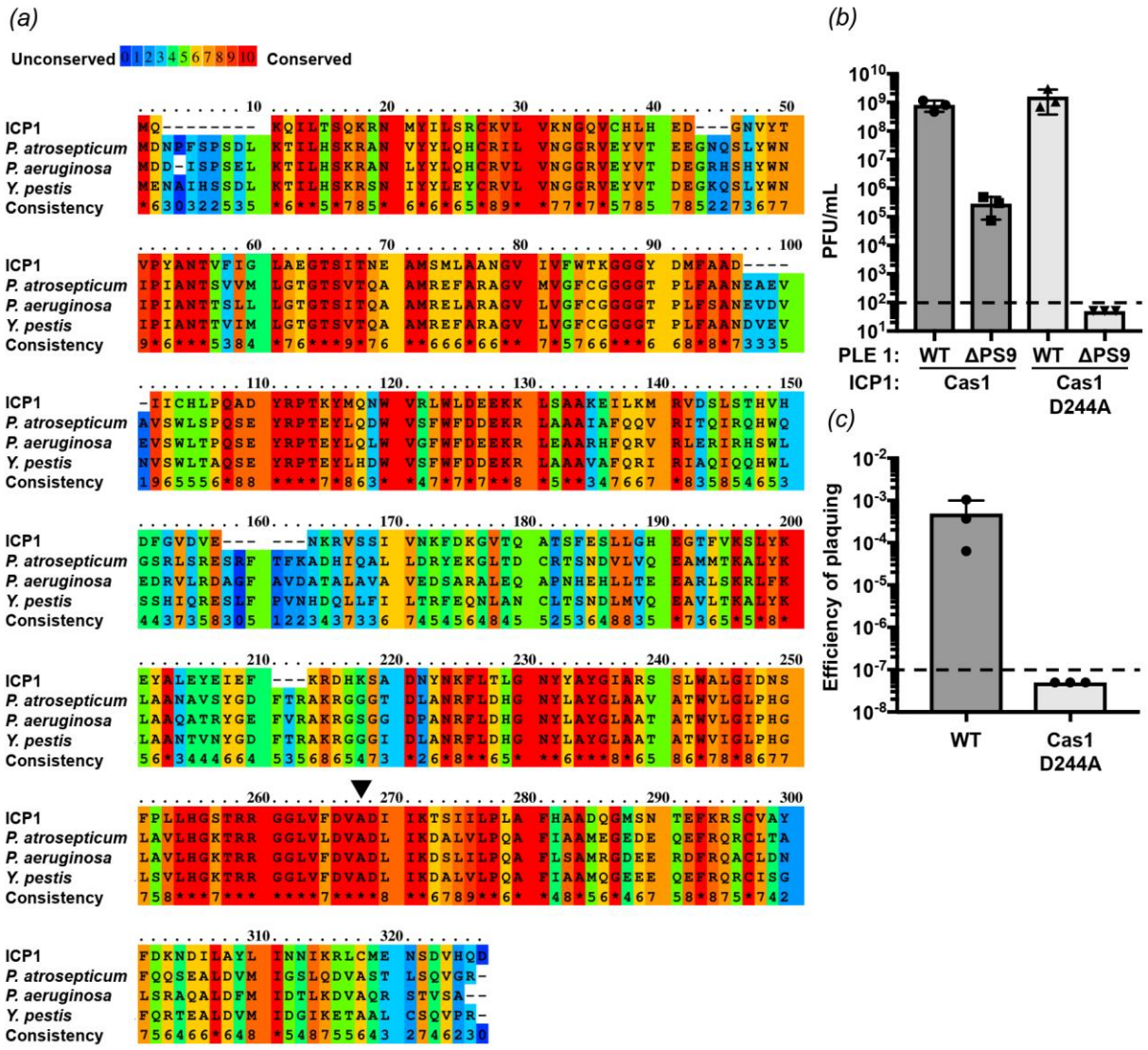
Supplementary Figure 3.3. High-throughput spacer acquisition mapping on circularized PLE 1. Spacers which uniquely mapped to PLE 1 on the plus or minus strand are indicated in red and blue, respectively. Black scale-bars at each quadrant represent a height of 50,000 spacer hits. The PLE 1 integration location is at the 12 o'clock scale-bar position. Putative open reading frames are indicated by white arrows. This representative graph displays results from replicate 2 in Supplementary Fig. 4.



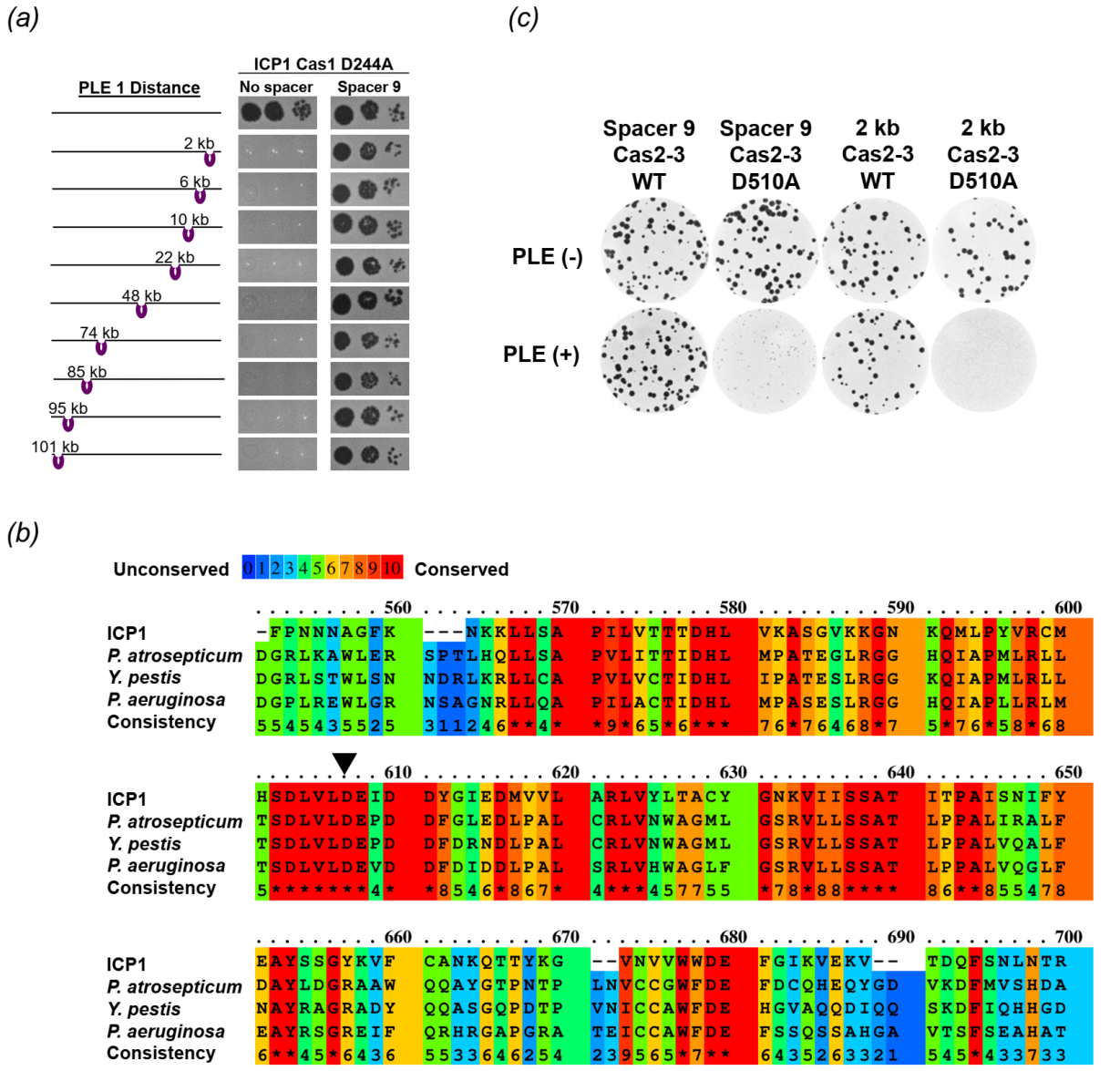
Supplementary Figure 3.4. Replicates of high-throughput ICP1 spacer acquisition mapping to the small chromosome. Spacer locations of the most leader-proximal spacers on the plus and minus strand are indicated in red and blue, respectively. Uniquely mapped spacers are shown in solid blue or red, while translucent bars show mapping of spacers to all possible locations. The scale bars measure the number of mapped spacers, and the tick marks around the chromosome are in 18kb intervals. Replicate 2 is the same data as in Fig. 3a and b and Supplementary Fig. 3.



Supplementary Figure 3.5. Sequence logo of PAMs of natural ICP1 isolates. The PAM sequence of the ICP1-encoded CRISPR-Cas system. Alignment of flanking sequence of all known targets of spacers found in natural ICP1 isolates. Sequence logos were generated using WebLogo [57].



Supplementary Figure 3.6. Conserved residue in ICP1's Cas1 is necessary for spacer acquisition. *(a)* Praline alignment [58] of Cas1 from ICP1 and other Type I-F systems in the organism indicated. The black arrow indicates the conserved residue mutated in ICP1 (D244A). *(b,c)* Engineered phage with a PLE internal spacer (S9) +/- Cas1 D244A were passaged on a PLE (+) host. Ten plaques of each phage were picked to generate a phage stock, and each stock was re-plaqued onto a PLE (+) host and a PLE $1^{\Delta PS9}$ host (as plaque formation on the PLE $1^{\Delta PS9}$ host requires spacer acquisition) and the number of plaques were counted *(b)*. The efficiency of plaquing *(c)* is calculated as the number of plaques on a PLE $^{\Delta PS9}$ host relative to the number of plaques on a WT PLE (+) host. Plaque formation on the PLE $1^{\Delta PS9}$ could not be detected in the Cas 1 D244A mutant (the dashed line indicates the limit of detection), compared to the WT phage in which approximately 1 phage per 2,835 acquired a new spacer that enabled plaque formation on the PLE $1^{\Delta PS9}$ host (EOP =$3.3x10^{-4}$). Error bars indicate standard deviation of three independent replicates.



Supplementary Figure 3.7. PLE transduced in unique sites in *V. cholerae* chromosome can be overcome by ICP1-CRISPR mediated interference dependent upon Cas2-3 helicase activity. *(a)* Tenfold dilutions of ICP1 without any spacer or with a PLE targeting spacer (S9) spotted on lawns of *V. cholerae*. Transductants used are the same as in Figure 3.6c. *(b)* Praline alignment [58] of a region of Cas2-3 from ICP1 and other Type I-F systems in the organism indicated. The black arrow indicates the conserved residue mutated in ICP1 (D510A). *(c)* Plaque plates of different concentrations of ICP1 with the labeled spacer in the WT or helicase dead (D510A) Cas2-3 background. The concentration of the 2 kb Cas2-3 D510A phage is so high that *V. cholerae* is unable to form a healthy lawn, but no plaques were detected.

Supplementary Table 1.1. Strains used in this study

Bacterial strains	Description	Source
E7946	*V. cholerae* O1, El Tor biotype; SmR, used as PLE (-)	[121]
PLE 1	E7946 containing PLE 1 integrated into VCR between VCA0329 and VCA0330	[22]
PLE 1 *Δint*	PLE 1 with an in-frame Spec-frt cassette replacing Orf1	This study
Δint* P_{tac}-*int	PLE 1 *Δint* with expression cassette modified from Dalia et al., 2014[50] for int under control of the P_{tac} promoter and a riboswitch integrated into the *V. cholerae lacZ* locus (*ΔlacZ*)	This study
Δint* P_{tac}-*EV	PLE 1 *Δint* with a P_{tac} promoter and a riboswitch integrated into the *V. cholerae lacZ* locus (*ΔlacZ*)	This study
PLE 1 FLAG-Int	PLE 1 with FLAG-tag fused to N-terminus of endogenous Int, Kanamycin resistance cassette inserted downstream of *int*	This study
miniPLE	E7946 containing PLE 1 *int* and Kanamycin resistance cassette integrated in VCR between VCA0329 and VCA0330	This study
BTH1101	*E. coli* adenylate cyclase knockout (*Δcya*), BACTH expression host	Lab collection
E. coli BL21	used for protein expression and purification	Lab collection
PLE 1-Kan^R	PLE 1, kanamycin resistance cassette inserted downstream of ORF 23	[22]
PLE 2-Kan^R	E7946 containing PLE 2 interrupting VCA0581, Kanamycin resistance cassette inserted downstream of ORF 27	[22]
PLE 3-Kan^R	E7946 containing PLE 3 integrated in VCR in between VCA0415 and VCA0416, Kanamycin resistance cassette inserted downstream of ORF 27	[22]
PLE 4-Kan^R	E7946 containing PLE 4 integrated in VCR in between VCA0353 and VCA0354, Kanamycin resistance cassette inserted downstream of ORF 29	[22]
PLE 5-Kan^R	E7946 containing PLE 3 integrated in VCR in between VCA0407 and VCA0408, Kanamycin resistance cassette inserted downstream of ORF 29	[22]
E7946 *ΔlacZ::att reporter*	Constitutively expressed *lacZ* cloned from *E. coli* flanked by part of a VCR that PLE 1 integrates into betweenVCA0462 and VCA0463 (*attC*) and the entire region between PLE 1 ORF23 and Int when PLE is circularized *(attP*)	This study
PLE1 ΔORFs2-5	PLE 1 ORF2- ORF5 replaced with an in-frame Spec-frt cassette	This study
PLE1 ΔORFs7-14	PLE 1 ORF7- ORF14 replaced with an in-frame Spec-frt cassette	This study
PLE1 ΔORFs21-23	PLE 1 ORF15- ORF20 replaced with an in-frame Spec-frt cassette	This study
PLE1 ΔORFs15-20	PLE 1 ORF21- ORF23 replaced with an in-frame Spec-frt cassette	This study
KS441	E7946 containing PLE 1 integrated into VCR between VCA0362 and VCA0363, used for *in vitro* recombination templates	[22]

Supplementary Table 1.2. Primers used in this study

Primer	Sequence (5' - 3')	Application
Zac14	AGGGTTTGAGTGCGATTACG	qPCR FWD
Zac15	TGAGGTTTTACCACCTTTTGC	qPCR RV
KS364	CCGCTATCTTTCGAGGTAGC	circularization PCR FWD/in vitro *attP* template FWD/*attR* RV
KS365	GCTACTCTCCGTTAAATTCCG	circularization PCR RV
ACM415	CATGGCTTTGCGATGATGG	*in vitro attC* template FWD/*attR* FWD
ACM416	GGTGGTGCTTGGGATAACTC	*in vitro attC* template RV/*attL* RV
KS289	TTGGCGTTTACTAGATACTCGTC	*in vitro attP* template RV/*attL* template FWD
KS587	ATTCCGGGGATCCGTCGACC	amplify FRT-SpecR cassette
KS586	TGTAGGCTGGAGCTGCTTCG	amplify FRT-SpecR cassette
ACM153	CTGACATTGATTTCCCTCCG	PLE 1 Δint
KS307	GGTCGACGGATCCCCGGAATTGGCATATATATTCACACACGC	PLE 1 Δint
KS308	CGAAGCAGCTCCAGCCTACATAAGAAAAAGACCGCCCTATTG	PLE 1 Δint
KS369	CGTAACTAAATTGGTGGTGTGC	PLE 1 Δint
KS372	GGTGATTAATTGCTATACAAGTGG	PLE1 ΔORFs2-5
KS371	CGAAGCAGCTCCAGCCTACATAAAATAAACCGCCTCAATAGGG	PLE1 ΔORFs2-5
KS477	GGTCGACGGATCCCCGGAATACTACTCACTTTATATAGTTTTCCTTATGTTG	PLE1 ΔORFs2-5
ks457	CTGAAGATAATCTAACGATAGTTTATCTAACG	PLE1 ΔORFs2-5
KS322	AGCGGAGCTATTAAGTATGC	PLE1 ΔORFs7-14
KS319	GGTCGACGGATCCCCGGAATCATAAGGTTGGCTCCTCAATG	PLE1 ΔORFs7-14
KS403	CGAAGCAGCTCCAGCCTACAAATTGGCTCGACTTAATTTAA	PLE1 ΔORFs7-14
KS379	GGCTATATGTGCGTGTAATGC	PLE1 ΔORFs7-14
KS334	TACTCCCTTAGCAAGGTTGG	PLE1 ΔORFs15-20
KS332	GGTCGACGGATCCCCGGAATTGTTGGCATGTTCGTATTTCC	PLE1 ΔORFs15-20
KS302	CGAAGCAGCTCCAGCCTACATAAAGGCTAGAAAAATATGAACAAAG	PLE1 ΔORFs15-20
KS304	CCTGCGTTAACAGCTTCTGC	PLE1 ΔORFs15-20
KS495	GAAGATGGTGAGGCACTAGC	PLE1 ΔORFs21-23
KS493	GGTCGACGGATCCCCGGAATGATGTTCATATTTTCCTATCCTATCC	PLE1 ΔORFs21-23
ACM363	CGAAGCAGCTCCAGCCTACATAACTACTGGTTGCACAAGATTAACC	PLE1 ΔORFs21-23
ACM26	CTATGTGAACCAAAAGTTGAGCG	PLE1 ΔORFs21-23

Supplementary Table 1.2 (continued). Primers used in this study

Primer	Sequence (5' - 3')	Application
KS501	GGATCCCCAGCTTCGCGTCCGCGGTCTTTTTCTTAC GTGATTTTAAGC	miniPLE
KS327	CGGTGAACGCTCTCCTGAGTACAAGATTAACCATAT ATGAGCCG	miniPLE
KS324	GGACGCGAAGCTGGGGATCC	amplify KanR cassette
KS325	ACTCAGGAGAGCGTTCACCG	amplify KanR cassette
ACM417	GTTTGTCATCATCATCTTTATAATCCATATATATTCA CACACGCTTAATAAAAAG	PLE 1 FLAG-Int
ACM418	GATTATAAAGATGATGATGACAAACCAAAACTTTACA TATTTAGACGAGTATCTAG	PLE 1 FLAG-Int
ACM321	GGATCCCCAGCTTCGCGTCCTCGGTAGTATTAGGC TAACGCCCGCCTAAGGGGCTGGCAACGCATTAGCA CCAAACTCAAACACAACAACTGCAACGCGCAACGC AATTAATGTGAGTTAGCTCACTCATTAGGCACCCCA GGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTG TGGAATTGTGAGCGGATAACAATTTCACACAGGATC CCGGGAGGAGGTAACGTAATGACCATGATTACGGA TTCACTGG	E7946 *ΔlacZ::att* reporter - gBlock containing *attC* and constitutive promoter
KS853	ATGACCATGATTACGGATTCAC	E7946 *ΔlacZ::att* reporter
KS555	TTATTTTTGACACCAGACCAACTGG	E7946 *ΔlacZ::att* reporter
KS565	CCAGTTGGTCTGGTGTCAAAAATAACTACTGGTTGC ACAAGATTAACC	E7946 *ΔlacZ::att* reporter
ACM343	GGTGTAGGCTGGAGCTGCTTCGGAGAGGCTTAACG CCTCTC	E7946 *ΔlacZ::att* reporter
ACM344	GAAGCAGCTCCAGCCTACACCACAATAAGCCAGAG AGCCTTAAG	E7946 *ΔlacZ::att* reporter
ACM316	CCTGCTAAGGAGGTAACAACAAGATGCCAAAACTTT ACATATTTAGACG	Δint P_{tac}-int
ACM317	CTCTCATCCGCCAAAACAGCTTACGTGATTTTAAGC TTGCGG	Δint P_{tac}-int
ACM292	TTTCGTCTTATGTGATCGATCGAGATATTGTGGTGA TG	ICP1 *ΔpexA*
ACM293	AAGACATCACCACAATATCTCGATCGATCACATAAG AC	ICP1 *ΔpexA*
ACM301	GCTATGACCATGATTACGCCACTAGACAATCTCACA AAAGAAG	ICP1 *ΔpexA*
ACM289	CGACGTTGTAAAACGACGGCCAAATAAGTTTGATGG CTGG	ICP1 *ΔpexA*
ACM290	GCAAGTTTTGGAGAATAATAATTTTAGAGGAAGGTG TGTATAATGAG	ICP1 *ΔpexA*
ACM291	GCAAGTTTTGGAGAATAATAATTTTAGAGGAAGGTG TGTATAATGAG	ICP1 *ΔpexA*

Supplementary Table 1.3. Plasmids used in this study

Plasmids	Description	Source
P_{tac}-*pexA*	pMMB67E plasmid engineered to contain a riboswitch downstream of P_{tac} for inducible expression of PexA	This study
P_{tac}-*EV*	pMMB67E plasmid engineered to contain a riboswitch downstream of P_{tac}, empty vector control	This study
P_{bad}-*orf49*	pBAD derivative plasmid engineered to contain a riboswitch downstream of P_{bad} for inducible expression of ICP1 ORF49	This study
P_{bad}-*orf50*	pBAD derivative plasmid engineered to contain a riboswitch downstream of P_{bad} for inducible expression of ICP1 ORF50	This study
P_{bad}-*orf51*	pBAD derivative plasmid engineered to contain a riboswitch downstream of P_{bad} for inducible expression of ICP1 ORF51	This study
P_{bad}-*orf52*	pBAD derivative plasmid engineered to contain a riboswitch downstream of P_{bad} for inducible expression of ICP1 ORF52	This study
pE-SUMO-*int*	Vector to express 6xHisSumo-fusion protein, fused to N-terminus of Int	This study
pE-SUMO-*pexA*	Vector to express 6xHisSumo-fusion protein, fused to N-terminus of PexA	This study
T18-Int	pUT18, T18 subunit of *cya* fused to C-terminus of *int*	This study
T25-PexA	pKNT25, T25 subunit of *cya* fused to C-terminus of *pexA*	This study
T18-EV	pUT18, T18 subunit of *cya*	Lab collection
T25-EV	pKNT25, T25 subunit of *cya*	Lab collection

Supplementary Table 1.4. Phage isolates used in this study

Phage isolates	Description	Source
ICP1	ICP1_2006_E *ΔCRISPR Δcas2-3*	This study
ICP1 *ΔpexA*	ICP1_2006_E *ΔCRISPR Δcas2-3 ΔpexA*	This study
ICP1 CRISPR+	ICP1_2006_E	[24]
$ICP1_{2004}$	ICP1_2004_A *Δcas2-3*	This study
$ICP1_{2004}$ *Δorf49-52*	ICP1_2004_A *Δcas2-3 ΔORFs49-52*	This study
$ICP1_{2004}$ ΔpexA	ICP1_2004_A *Δcas2-3 ΔpexA*	This study
$ICP1_{2005}$	ICP1_2005_A *ΔCRISPR Δcas2-3*	This study
$ICP1_{2005}$ *ΔpexA*	ICP1_2005_A *ΔCRISPR Δcas2-3 ΔpexA*	This study
$ICP1_{2011}$	ICP1_2011_A *ΔCRISPR Δcas2-3*	This study
$ICP1_{2011}$ *ΔpexA*	ICP1_2011_A *ΔCRISPR Δcas2-3 ΔpexA*	This study

Supplementary Table 1.5. Phage isolates used in PexA sequence analysis

Phage Isolate	Year	Accession	Reference
ICP1_2001_A	2001	HQ641347	[19]
ICP1_2004_A	2004	HQ641354	[19]
ICP1_2005_A	2005	HQ641352	[19]
ICP1_2006_A	2006	HQ641351	[19]
ICP1_2006_B	2006	HQ641350	[19]
ICP1_2006_C	2006	HQ641349	[19]
ICP1_2006_D	2006	HQ641348	[19]
ICP1_2006_E	2006	MH310934	[21]
ICP1_2011_A	2011	MH310933	[21]
JSF01	2001	KY883636	[25]
JSF02	2001	KY883637	[25]
JSF04	2001	KY065147	[25]
JSF05	2002	KY883634	[25]
JSF06	2002	KY883635	[25]
JSF13	2009	KY883638	[25]
JSF14	2011	KY883639	[25]
JSF17	2012	KY883640	[25]

Supplementary Table 2.1. Escape phage gene products that acquired mutations

Phage	*gp63*	*gp121*	*gp139*	*gp147*	*gp176*
escape Φ1				Q211E	
escape Φ2			M66I	Δ16 bp	
escape Φ3	R12C	I321I		Δ16 bp	A274T

Supplementary Table 2.2 Strains used in this study

Bacterial strains	Strain number	Description	Source
PLE (-)	KDS6	*V. cholerae* O1, El Tor biotype; SmR, E7946	[121]
PLE	KDS 36	E7946 containing PLE 1 integrated into VCR between VCA0329 and VCA0330	[22]
P_{orf2}-*nanoluc*	KS2119	PLE with nanoluciferase and KanR cloned downstream of *orf2*	This study
miniPLE	ACM270	E7946 containing PLE 1 *int* and Kanamycin resistance cassette integrated in VCR between VCA0329 and VCA0330	[29]
miniPLE$_{CD}$	ACM296	miniPLE *int S11A*	This study
PLE 2	KDS37	E7946 containing PLE 2 interrupting VCA0581	[22]
PLE 3	KDS38	E7946 containing PLE 3 integrated in VCR in between VCA0415 and VCA0416	[22]
PLE 4	KDS39	E7946 containing PLE 4 integrated in VCR in between VCA0353 and VCA0354	[22]
PLE 5	KDS40	E7946 containing PLE 3 integrated in VCR in between VCA0407 and VCA0408	[22]

Supplementary Table 2.3 Plasmids used in this study

Plasmids	Description	Source
P_{tac}-*EV*	pMMB67E plasmid engineered to contain a riboswitch downstream of P_{tac} empty vector control	This study
P_{tac}-*helA*	pMMB67E plasmid engineered to contain a riboswitch downstream of P_{tac}, for inducible expression of ICP1^{A} *helA*	This study
P_{tac}-*helB*	pMMB67E plasmid engineered to contain a riboswitch downstream of P_{tac}, for inducible expression of ICP1^{B} *helB*	This study
P_{tac}-*dda*	pMMB67E plasmid engineered to contain a riboswitch downstream of P_{tac}, for inducible expression of T4 *dda*	This study

Supplementary Table 2.4 Primers used in this study

Primer	Sequence (5' - 3')	Application
Zac14	AGGGTTTGAGTGCGATTACG	PLE qPCR FWD
Zac15	TGAGGTTTTACCACCTTTTGC	PLE qPCR RV
Zac68	CTGAATCGCCCTACCCGTAC	ICP1 qPCR FWD
Zac69	GTGAACCAACCTTTGTCGCC	ICP1 qPCR RV
Zac10	ATGCAATGCAGCCATAAACA	miniPLE qPCR FWD
Zac11	GCGTTTAGTTCGGTGTTGGT	miniPLE qPCR RV
KS364	CCGCTATCTTTCGAGGTAGC	circularization PCR FWD
KS365	GCTACTCTCCGTTAAATTCCG	circularization PCR RV
KS587	ATTCCGGGGATCCGTCGACC	amplify nanoluciferase
KS586	TGTAGGCTGGAGCTGCTTCG	amplify nanoluciferase
KS324	GGACGCGAAGCTGGGGATCC	Amplify kanR cassette
KS325	ACTCAGGAGAGCGTTCACCG	Amplify kanR cassette
KS372	GGTGATTAATTGCTATACAAGTGG	P_{orf2}-*nanoluc*
KS501	GGATCCCCAGCTTCGCGTCCGCGGTCTTTTTCTTACGTGATTTTAAGC	P_{orf2}-*nanoluc*
KS502	CGGTGAACGCTCTCCTGAGTCCTATTGAGGCGGTTTATTTTATGTG	P_{orf2}-*nanoluc*
KS1168	CGAACGCATTCTGGCGTAAACACATAAAATAAACCGCCTCAATAGG	P_{orf2}-*nanoluc*
KS1169	GTGAAGACCATTACGTTACCTCCttttatgtgtTTAGGAAACCGTAGACAG	P_{orf2}-*nanoluc*
KS369	CGTAACTAAATTGGTGGTGTGC	P_{orf2}-*nanoluc*
ACM153	CTGACATTGATTTCCCTCCG	miniPLE$_{CD}$
ACM219	GGCGTTTACTAGCTACTCGTC	miniPLE$_{CD}$
ACM218	GACGAGTAGCTAGTAAACGCC	miniPLE$_{CD}$
ACM26	CTATGTGAACCAAAAGTTGAGCG	miniPLE$_{CD}$
gp58F	AACGCTGCTTTTCCTTTTGA	ICP1 *gp58* detection
gp58R	CCCAGCATTGAGGACACTTT	ICP1 *gp58* detection
ACM295	TTGCACTTGCAGCAACATGG	ICP1 *helA* detection
ACM367	AAAGCGTTCAATACGACGCC	ICP1 *helA* detection
ACM531	GTTTCTACTACTGTACCGAC	ICP1 *helB* detection
ACM583	TGGTAATCCTATCCCTGATG	ICP1 *helB* detection

Supplementary Table 2.5 Phages used in this study

Phage isolates	Strain number	Description	Source
ICP1^{A}	KSΦ38	ICP1_2006_E *ΔCRISPR Δcas2-3*	[29]
ICP1^{A} *ΔpexA*	ACMΦ142	ICP1_2006_E *ΔCRISPR Δcas2-3 ΔpexA*	[29]
ICP1^{A} *ΔhelA*	ACMΦ262	ICP1_2006_E *ΔCRISPR Δcas2-3 ΔhelA*	This study
ICP1^{A} *ΔpexA ΔhelA*	ACMΦ264	ICP1_2006_E *ΔCRISPR Δcas2-3 ΔpexA ΔhelA*	This study
ICP1^{B}	ACMΦ259	ICP1_2017_A_Mathbaria *Δcas2-3;* ICP1 isolate recovered from cholera patient stool collected from ICDDR,B and engineered to be *Δcas2-3*	This study
ICP1^{B} *ΔhelB*	ACMΦ262	ICP1_2017_A_Mathbaria *Δcas2-3 ΔhelB*	This study
ICP1^{B} B *ΔHD*	ACMΦ288	ICP1_2017_A_Mathbaria *Δcas2-3 Δ118-192*	This study

Supplementary Table 2.6 Phage isolates used in HelA sequence analysis

Phage	Accession	Source
ICP1	HQ641347	[19]
ICP1_2001_A	HQ641353	[19]
ICP1_2001_B	KY883636	[25]
ICP1_2001_C	KY883637	[25]
ICP1_2001_D	KY065147	[25]
ICP1_2001_E	KY883634	[25]
ICP1_2001_F	KY883635	[25]
ICP1_2004_A	HQ641354	[19]
ICP1_2005_A	HQ641352	[21]
ICP1_2006_A	HQ641351	[19]
ICP1_2006_B	HQ641350	[19]
ICP1_2006_E	MH310934	[21]
ICP1_2009_A	KY883638	[25]
ICP1_2011_A	MH310933	[21]
ICP1_2011_B	MH310935	[21]
ICP1_2011_C	KY883639	[25]
ICP1_2012_B	MH310936	[25]
ICP1_2015_A	*	[69]
ICP1_2016_A	*	[69]
ICP1_2017_A	*	[69]
ICP1_2017_B	*	[69]
ICP1_2017_C	*	[69]
ICP1_2017_D	*	[69]
ICP1_2017_E	*	[69]
ICP1_2017_F	*	[69]

*Sequenced genomes not available. *helA* was amplified and sequenced using Sanger Sequencing.

Supplementary Table 2.7 Proteins used in SF1B helicase phylogeny

Phage isolate	Accession	Reference
ICP1^{A} HelA	ADX89559	[19]
ICP1^{B} HelB	ADX88195	[19]
***Pseudoalteromonas* phage^{1}**	APC44385	[122]
***Pseudoalteromonas* phage^{2}**	ASU03327	
***Pseudoalteromonas* phage^{3}**	YP_009225645	
***Shewanella* phage^{1}**	YP_009104072	[123]
***Shewanella* phage^{2}**	YP_009100375	[123]
T4 Dda	ADJ39734	[124]
Vibrio phage^{1}	AUR92300	[125]
Vibrio phage^{2}	AUR89265	[125]
Vibrio phage^{3}	AUR91655	[125]
Vibrio phage^{4}	YP_007877404	
Vibrio phage^{5}	AUR93419	[125]
Vibrio phage^{6}	AUR86306	[125]
Vibrio phage^{7}	YP_007673653	
Vibrio phage^{8}	AUR87502	[125]
Vibrio phage^{9}	AOQ26819	
Vibrio phage^{10}	AUR8488	[125]
Vibrio phage^{11}	YP_007676005	
Vibrio phage^{12}	AUR94225	[125]
Vibrio phage^{13}	BAV80879	[126]

Supplementary Table 3.1: Phage used in this study

Phage	Description	Source
ICP1_2011_A	CRISPR-Cas positive ICP1 isolate recovered from cholera patient stool with KS393	[24]
ICP1_2001_E	CRISPR-Cas positive ICP1 isolated collected from water sample in Bangladesh, referred to as JSF5 in original publication	[25]
ICP1_2001_F	CRISPR-Cas positive ICP1 isolated collected from water sample in Bangladesh, referred to as JSF6 in original publication	[25]
ICP1_2005_A	CRISPR-Cas positive ICP1 isolate from cholerae patient stool collected at ICDDR,B	[19]
ICP1_2006_E	CRISPR-Cas positive ICP1 isolate from cholerae patient stool collected at ICDDR,B	[21]
ICP1_2004_A	CRISPR-Cas positive ICP1 isolate from cholerae patient stool collected at ICDDR,B	[19]
ICP1_2009_A	CRISPR-Cas positive ICP1 isolated collected from water sample in Bangladesh, referred to as JSF13 in original publication	[25]
ICP1_2011_B	CRISPR-Cas positive ICP1 genome assembled from metagenomic analysis of diarrheal stool sample from ICDDR,B	[21]
ICP1_2011_C	CRISPR-Cas positive ICP1 isolated collected from water sample in Bangladesh, referred to as JSF14 in original publication	[25]
ICP1_2012_B	CRISPR-Cas positive ICP1 isolated collected from water sample in Bangladesh, referred to as JSF17 in original publication	[25]
ICP1_2015_A	CRISPR-Cas positive ICP1 isolate recovered from cholera patient stool collected from ICDDR,B	This study
ICP1_2016_A	CRISPR-Cas positive ICP1 isolate recovered from cholera patient stool collected from ICDDR,B	This study
ICP1_2017_A	CRISPR-Cas positive ICP1 isolate recovered from cholera patient stool collected from ICDDR,B	This study
ICP1_2017_B	CRISPR-Cas positive ICP1 isolate recovered from cholera patient stool collected from ICDDR,B	This study
ICP1_2017_C	CRISPR-Cas positive ICP1 isolate recovered from cholera patient stool collected from ICDDR,B	This study
ICP1_2017_D	CRISPR-Cas positive ICP1 isolate recovered from cholera patient stool collected from ICDDR,B	This study
ICP1_2017_E	CRISPR-Cas positive ICP1 isolate recovered from cholera patient stool collected from ICDDR,B	This study
ICP1_2017_F	CRISPR-Cas positive ICP1 isolate recovered from cholera patient stool collected from ICDDR,B	This study
ICP1_2011_A ΔS9	Spontaneous deletion of spacer 9 in phage CRISPR array	[24]
ICP1_2011_A S10	ICP1_2011_A with expanded CRISPR array with new spacer against PLE in leader proximal CRISPR 1 array, isolated from selection of ICP1_2011_A on PLE PS8* PS9* host	This study
ICP1_2011_A ΔS2-9	Directed deletion of spacers 2-9 in phage CRISPR array	[53]
ICP1_2011_A ΔS9, new mu spacer	ICP1_2011_A ΔS9 with new spacer against mu-like region in leader proximal CRISPR 1 array, isolated from high-throughput spacer acquisition pool	This study

ICP1_2011_A ΔS2-9, Cas1D244A	Engineered Cas1D244 mutation to prevent spacer acquisition in ICP1_2011_A	This study
ICP1_2011_A -93kb (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study
ICP1_2011_A -71kb (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study
ICP1_2011_A -53kb (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study
ICP1_2011_A -46kb (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study
ICP1_2011_A -22kb (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study
ICP1_2011_A -2.5kb (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study
ICP1_2011_A -1.5kb (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study
ICP1_2011_A PLE Internal (S9) (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study
ICP1_2011_A 0.5kb (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study
ICP1_2011_A 22.5kb (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study
ICP1_2011_A 437kb (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study
ICP1_2011_A GA PAM (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study
ICP1_2011_A GG (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study

ICP1_2011_A GT (+/- Cas1D244A)	ICP1_2011_A ΔS2-9 engineered with pCRISPR editing template for custom spacer selected from acquisition pool at a specific distance from PLE 1. Constructed in ΔS2-9 and ΔS2-9 Cas1D244A background	This study

Supplementary Table 3.2: Bacterial strains used in this study

Strain	Description	Source
KS393	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	[24]
KDS2	KS393 PLE 1 PS8* such that target of ICP1_2011 CRISPR spacer 8 target is mutated	[24]
KDS3	KS393 PLE 1 PS8*, Kanamycin resistance cassette inserted downstream of ORF 23	[22]
KDS4	KS393 PLE 1 Kanamycin resistance cassette inserted downstream of ORF 23	[22]
KS1265	KS393 *ΔPLE 1::Spec*. Natural transformation used to knockout PLE 1 and replace it with a spectinomycin cassette	This study
KDS46	KDS3 PLE 1 *Δorf3*. Natural transformation used to knockout PLE 1 orf3 and replace it with a frt scar	This study
ACM71	KDS2 *Δmu*. Natural transformation used to knockout mu-like region in large chromosome and replace it with a kanamycin cassette	This study
ACM146	KDS2 *Δmu, Δmu2*. Natural transformation used to knockout mu-like region in small chromosome in KS393 Δmu background and replace it with a spectinomycin cassette	This study
ACM584	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	This study
ACM585	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	This study
ACM589	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	This study
ACM592	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	This study
ACM596	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	This study
ACM598	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	This study
ACM604	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	This study
ACM606	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	This study
ACM608	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	This study
ACM612	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	This study
ACM614	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	This study
ACM618	*V. cholerae* clinical Isolate from the ICDDR, B, harbors PLE 1	This study
KS1559	*V. cholerae* clinical isolate, PLE (-), P*tac*-CRISPR-Cas from O395, KanR ΔlacZ:: SpecR, constructed as per Box *et al*, 2016 for engineering ICP1 phage using the *V. cholerae* classical biotype Type I-E CRISPR-Cas system	This study
KDS187	Transductant in KS1265 background harboring PLE 1-KanR integrated in VCR in between VCA498 and VCA499. KS1265 added to supernatants of KDS3 infected with ICP1_2011_A ΔS9, KanR SpecR recipient selected and PLE insertion site confirmed by sequencing	This study
KDS188	Transductant in KS1265 background harboring PLE 1-KanR integrated in VCR in between VCA491 and VCA492. KS1265 added to supernatants of KDS3 infected with ICP1_2011_A ΔS9, KanR SpecR recipient selected and PLE insertion site confirmed by sequencing	This study
KDS189	Transductant in KS1265 background harboring PLE 1-KanR integrated in VCR in between VCA483 and VCA484. KS1265 added to supernatants of KDS3 infected with ICP1_2011_A ΔS9, KanR SpecR recipient selected and PLE insertion site confirmed by sequencing	This study

KDS190	Transductant in KS1265 background harboring PLE 1-KanR integrated in VCR in between VCA440 and VCA441. KS1265 added to supernatants of KDS3 infected with ICP1_2011_A ΔS9, KanR SpecR recipient selected and PLE insertion site confirmed by sequencing	This study
KDS191	Transductant in KS1265 background harboring PLE 1-KanR integrated in VCR in VCA354. KS1265 added to supernatants of KDS3 infected with ICP1_2011_A ΔS9, KanR SpecR recipient selected and PLE insertion site confirmed by sequencing	This study
KDS192	Transductant in KS1265 background harboring PLE 1-KanR integrated in VCR in between VCA336 and VCA337. KS1265 added to supernatants of KDS3 infected with ICP1_2011_A ΔS9, KanR SpecR recipient selected and PLE insertion site confirmed by sequencing	This study
KDS193	Transductant in KS1265 background harboring PLE 1-KanR integrated in VCR in between VCA312 and VCA313. KS1265 added to supernatants of KDS3 infected with ICP1_2011_A ΔS9, KanR SpecR recipient selected and PLE insertion site confirmed by sequencing	This study
KDS194	Transductant in KS1265 background harboring PLE 1-KanR integrated in VCR in between VCA301 and VCA302. KS1265 added to supernatants of KDS3 infected with ICP1_2011_A ΔS9, KanR SpecR recipient selected and PLE insertion site confirmed by sequencing	This study

Supplementary Table 3.3: Plasmids used in this study

Plasmid	Description	Source
pCRISPR	Derived from pMMB67EH, harbors *V. cholerae* CRISPR array under control of IPTG inducible promoter and insertion site for editing templates, AmpR	[53]
PTargetΦCR1S1-Ed ICP1 CRISPR array	pCRISPR cloned with anti-ICP1_2011_A S1 spacers and synthetic ICP1_2011_A CRISPR array with up and down arms of homology to permit recombination of synthetic array in phage genome. The synthetic phage CRISPR array harbors a BsaI site flanked by the phage direct repeats. Primers ACM485-512 with BsaI compatible arms were used to insert desired spacer sequences into this plasmid to render ICP1_2011_A with an engineered spacer	This study
pTargetΦ*cas1*-EdCas1D244A	pCRISPR cloned with anti-*cas1* spacer directed against ICP1_2011_A, and an editing template to introduce D244A point mutation	This study

Supplementary Table 3.4: Primers used in this study

	Primer name	Sequence (5' to 3' orientation)	Source
HTML_PCR of Leader proximal spacer in CRISPR1	KS395	TCTCTAAATTTTTAATAAACCTTTGAAATTTATGGAT	This study
	OLJ376	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGGGGGGGGGGGG	[127]

	KS397	AATGATACGGCGACCACCGAGATCTACACTCTTTG AAATTTATGGATTTTGTGGTATAGTTTAAGTG	This study
	BC43*	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCT	117
	BC44*	CAAGCAGAAGACGGCATACGAGATTACAAGGTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCT	117
	BC45*	CAAGCAGAAGACGGCATACGAGATTGTTGACTGTG ACTGGAGTTCAGACGTGTGCTCTTCCGATCT	117
Illumina sequencing primer	KS396	GGATTTTGTGGTATAGTTTAAGTGTTAGCAGCCGC	This study
qPCR for PLE replication	Zac14	AGGGTTTGAGTGCGATTACG	22
	Zac15	TGAGGTTTTACCACCTTTTGC	22
ICP1 gp58 detection	gp58F	AACGCTGCTTTTCCTTTTGA	19
	gp58R	CCCAGCATTGAGGACACTTT	19
ICP1 CRISPR 1 detection	KS252	CGGAGAAATTCAACAGTTATGG	24
	KS253	CAGAAAGTATTGCGGCTAGG	24
PLE detection	ACM349	GGTGATTGTCCTTTCATTTGGG	This study
	ACM350	GTGATGCTGTTCAAACAGGG	This study
	KS1034	TGCATATCCATGCACAGGTG	This study
	KS1035	TGAGGTTTTACCACCTTTTGC	This study
	KS1036	GGAACAAACACAACAACGC	This study
	KS1037	TGTGTTGTTATGTTTGGTGC	This study
	KS1038	TTACATGTTGGAACTGGTCG	This study
	KS1039	CAGCGGCCAAAAATTCTTCC	This study
Construction of custom ICP1_2011_A CRISPR spacer array[a]	ACM484 (-2.5 kb)**	CTAAC**AAGCCGAGAAAATAAAGGTGCTTGCTGGCA TA**	This study
	ACM485 (-2.5 kb)**	AAGA**TATGCCAGCAAGCACCTTTATTTTCTCGGCT T**G	This study
	ACM486 (-1.5 kb)**	CTAAC**CAAAATTCTTCCTATAGTCTGCTGTATAAC CA**	This study
	ACM487 (-1.5 kb)**	AAGA**TGGTTATACAGCAGACTATAGGAAGAATTTT G**G	This study
	ACM488 (22.5 kb)**	CTAAC**GTGGTTGTCCTTTCTTCTGGTTCAAGCAGC AT**	This study
	ACM489 (22.5 kb)**	AAGA**ATGCTGCTTGAACCAGAAGAAAGGACAACCA C**G	This study
	ACM490 (GG PAM)**	CTAAC**ATCCTGTTCGGTTCAATTGTCGAAGATGGT GA**	This study
	ACM491 (GG PAM)**	AAGA**TCACCATCTTCGACAATTGAACCGAACAGGA T**G	This study
	ACM492 (GT PAM)**	CTAAC**TCAACAGGTCGAGTTGATGAAAATTTTAGC TA**	This study
	ACM493 (GT PAM)**	AAGA**TAGCTAAAATTTTCATCAACTCGACCTGTTG A**G	This study
	ACM495 (-93 kb)**	CTAAC**ATCTTGAGAAACGAGTGTGGACAGTCCCCG AA**	This study
	ACM496 (-93 kb)**	AAGA**TTCGGGGACTGTCCACACTCGTTTCTCAAGA T**G	This study

	ACM497 (-71 kb)**	CTAAC**AGTCCCAATGAGCCGCGGTGGTTTCGGTTGTT**	This study
	ACM498 (-71 kb)**	AAGA**AACAACCGAAACCACCGCGGCTCATTGGGAC**T**G**	This study
	ACM499 (-53 kb)**	CTAAC**ATTGAGCCGAGGTGGTTTCGGTTGTGGTGTTT**	This study
	ACM500 (-53 kb)**	AAGA**AAACACCACAACCGAAACCACCTCGGCTCAA**T**G**	This study
	ACM501 (-46 kb)**	CTAAC**ATTGAGCCGCAGTGGTTACGGTTATTGTGTTT**	This study
	ACM502 (-46 kb)**	AAGA**AAACACAATAACCGTAACCACTGCGGCTCAA**T**G**	This study
	ACM503 (-22 kb)**	CTAAC**CATATTGTCGCGAATGTCGAGGCTTGCATTGT**	This study
	ACM504 (-22 kb)**	AAGA**ACAATGCAAGCCTCGACATTCGCGACAATAT**G**G**	This study
	ACM505 (0.5 kb)**	CTAAC**GAAGTTGGGATTTTGATGATTTTAGCCTTTAA**	This study
	ACM506 (0.5 kb)**	AAGA**TTAAAGGCTAAAATCATCAAAATCCCAACTT**C**G**	This study
	ACM509 (437 kb)**	CTAAC**TTTCGGTGGAGCAGTATCAAAATAGCGTTGTC**	This study
	ACM510 (437 kb)**	AAGA**GACAACGCTATTTTGATACTGCTCCACCGAA**A**G**	This study
Deletion of mu-like region	KS802	GGATCCCCAGCTTCGCGTCCCGCTACAAATAGTTTGGATCTCG	This study
	KS803	CGGTGAACGCTCTCCTGAGTTCACAAGGGAGTAGAAGCGG	This study
	KS804	ATGTTGCATCAGGGAACTGG	This study
	KS805	CAATGATAAAAGCAGCATGGG	This study
Deletion of mu2-like region	ACM103	CACCGATGGAAATCTGTGG	This study
	ACM104	GGATCCCCAGCTTCGCGTCCGCCGACAATGTTAATGTAAGTGTG	This study
	ACM105	GCCTAACTATGGTGTGGTCG	This study
Construction of Cas1 D244A	ACM471***	AAGAC**GTGGTGGGTTGGTTTTCGATGTAGCTGACATCA**	This study
	ACM472***	TTTC**TGATGTCAGCTACATCGAAAACCAACCCACCAC**G	This study
	ACM478	ATTATTGAAGTTTTGATGATAGCAGCTACATCGAAAACCAACCC	This study
	ACM479	TGCTATCATCAAAACTTCAATAATCTTACCTTTAGCTTTTCACGCTGC	This study

*The 6 bp barcode is underlined.

[a]All oligos harbor BsaI compatible overhangs for insertion into pTarget□CR1S1 -Ed ICP1 CRISPR array. Target as in manuscript is in parentheses

**The 32bp spacer sequence is in bold

***The 33bp spacer sequence is in bold

Supplementary Table 3.5: The majority of spacers in the ICP1-related phages CRISPR arrays show identity to *V. cholerae* PLEs

Phage	CRISPR	Spacer	Sequence	Match (%)[a]
ICP1_2001_E	CR 1	1	TGTGTCTATACTCAACCAATTTAAGCGCCGCA	ICP1 (81.25%)
		2	CTACTCTCCCCAATATTAGCCATTCCTAATTCA	-
		3	GTCACCTTACCGTAAGACAGGCAGTAAAATTA	PLE 2 (100%)
		4	AAACTAGTGGACGTAATGCAGTATTCACGGTT	PLE 1 (100%) PLE 2 (90.62%) PLE 4 (100%) PLE 5 (100%)
	CR 2	1	ATCCACACTACAAATAGAACACTCAACCGTGA	PLE 1 (100%) PLE 2 (87.5%) PLE 3 (81.25%) PLE 4 (100%) PLE 5 (100%)
		2	TATTGATTGGTCTCTAACCTTGGGATGATTAA	-
		3	AGCGTGTGGGCTTTCATTTTTAAGCCAGTAAA	PLE 1 (90.62%) PLE 2 (90.62%) PLE 3 (90.62%) PLE 4 (87.5%) PLE 5 (100%)
		4	TTCACGGGTAGCAACAGGGTAATAAACCAATA	PLE 1 (100%) PLE 2 (100%) PLE 4 (93.75%) PLE 5 (100%)
ICP1_2001_F	CR 1	1	TGTGTCTATACTCAACCAATTTAAGCGCCGCA	ICP1 (81.25%)
		2	CTACTCTCCCCAATATTAGCCATTCCTAATTCA	-
		3	GTCACCTTACCGTAAGACAGGCAGTAAAATTA	PLE 2 (100%)
		4	AAACTAGTGGACGTAATGCAGTATTCACGGTT	PLE 1 (100%) PLE 2 (90.62%) PLE 4 (100%) PLE 5 (100%)
	CR 2	1	ATCCACACTACAAATAGAACACTCAACCGTGA	PLE 1 (100%) PLE 2 (87.5%) PLE 3 (81.25%) PLE 4 (100%) PLE 5 (100%)

		2	TATTGATTGGTCTCTAACCTTGGGATGATTAA	-
		3	AGCGTGTGGGCTTTCATTTTTAAGCCAGTAAA	PLE 1 (90.62%) PLE 2 (90.62%) PLE 3 (90.62%) PLE 4 (87.5%) PLE 5 (100%)
		4	TTCACGGGTAGCAACAGGGTAATAAACCAATA	PLE 1 (100%) PLE 2 (100%) PLE 4 (93.75%) PLE 5 (100%)
ICP1_2004_A	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TGTTAGAGTCGGTAGTATCTGGATGATCGATA	PLE 2 (100%) PLE 3 (100%) PLE 4 (100%)
		3	TAGAAGAGTAATAGGAGCTACTGCAAACTTGT	-
		4	TAACTATGTGTGGTTTATATTTTGTGTGCAAGA	-
		5	TTTTGAAACTATTGACAGAAGGTTGGGAACCT	-
		6	TTGAGGTTGAACCTCTTCCGGTTCCTCTTCTG	PLE 1 (93.75%) PLE 2 (100%) PLE 3 (84.38%)
	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
ICP1_2005_A	CR 1	1	TGTGTCTATACTCAACCAATTTAAGCGCCGCA	ICP1 (81.25%)
		2	CTACTCTCCCCAATATTAGCCATTCCTAATTCA	-
		3	AAACTAGTGGACGTAATGCAGTATTCACGGTT	PLE 1 (100%) PLE 2 (90.62%) PLE 4 (100%) PLE 5 (100%)
		4	ATAATCGTTTTGAGTCTCACCAGCTTTTAGGC	PLE 2 (100%) PLE 3 (100%)
	CR 2	1	ATCCACACTACAAATAGAACACTCAACCGTGA	PLE 1 (100%) PLE 2 (87.5%) PLE 3 (81.25%) PLE 4 (100%) PLE 5 (100%)
		2	TATTGATTGGTCTCTAACCTTGGGATGATTAA	-

		3	TTCACGGGTAGCAACAGGGTAATAAACCAATA	PLE 1 (100%) PLE 2 (100%) PLE 4 (93.75%) PLE 5 (100%)
ICP1_2006_E	CR1	1	TGTGTCTATACTCAACCAATTTAAGCGCCGCA	ICP1 (81.25%)
		2	CTACTCTCCCCAATATTAGCCATTCCTAATTCA	-
		3	GTCACCTTACCGTAAGACAGGCAGTAAAATTA	PLE 2 (100%)
		4	AAACTAGTGGACGTAATGCAGTATTCACGGTT	PLE 1 (100%) PLE 2 (90.62%) PLE 4 (100%) PLE 5 (100%)
	CR 2	1	ATCCACACTACAAATAGAACACTCAACCGTGA	PLE 1 (100%) PLE 2 (87.5%) PLE 3 (81.25%) PLE 4 (100%) PLE 5 (100%)
ICP1_2009_A	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TGTTAGAGTCGGTAGTATCTGGATGATCGATA	PLE 2 (100%) PLE 3 (100%) PLE 4 (100%)
		3	TCACAATCAGCTATAAGCCCTGCATTTTCAAT	PLE 2 (100%)
		4	TTGTAGTGATGACATAATCTCGTCTCGACTCA	PLE 2 (100%)
		5	AAGCAGAACTCACCGCCGAAGTGGAACAGCGT	
	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
		2	AAGACGTGACAGCAGTGATCGACTTTATAACA	PLE 2 (100%)
ICP1_2011_A	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TGTTAGAGTCGGTAGTATCTGGATGATCGATA	PLE 2 (100%) PLE 3 (100%) PLE 4 (100%)
		3	TTATGTATTGACCCCGACACGCCCCCCGACTG	-
		4	TTACAGACGACCTAACTCTTCAGTACCATGAT	-
		5	TACATAAGCTGCAACACGGTGTTCGTTTAAGT	PLE 2 (96.88%) PLE 4 (96.88%)

		6	AAAATACGCCTTTTTCCCTTCATCGTTTAAAG	PLE 3 (90.62%) PLE 4 (100%)
		7	ACCAACAAATCCCATAAACTGATAACCACGTT	-
		8	GTCAACCCTTTGCTTATCTTCCCTATTTAAAT	PLE 1 (100%) PLE 2 (100%)
		9	TGTTAACCACCGCTTGAAATAATCATGATGCA	PLE 1 (100%) PLE 2 (100%)
ICP1_2011_B	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TGTTAGAGTCGGTAGTATCTGGATGATCGATA	PLE 2 (100%) PLE 3 (100%) PLE 4 (100%)
		3	TCACAATCAGCTATAAGCCCTGCATTTTCAAT	PLE 2 (100%)
		4	TTGTAGTGATGACATAATCTCGTCTCGACTCA	PLE 2 (100%)
		5	AAGCAGAACTCACCGCCGAAGTGGAACAGCGT	PLE 1 (100%) PLE 2 (100%) PLE 3 (84.38%) PLE 5 (81.25%)
		6	TTGCATCAGTTGGATAGTTAATTGAGTGGGGC	PLE 1 (100%) PLE 4 (93.75%) PLE 5 (100%)
	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
		2	AAGACGTGACAGCAGTGATCGACTTTATAACA	PLE 2 (100%)
ICP1_2011_C	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TCACAATCAGCTATAAGCCCTGCATTTTCAAT	PLE 2 (100%)
		3	TTGTAGTGATGACATAATCTCGTCTCGACTCA	PLE 2 (100%)
		4	AAGCAGAACTCACCGCCGAAGTGGAACAGCGT	PLE 1 (100%) PLE 2 (100%) PLE 3 (84.38%) PLE 5 (81.25%)
		5	TTGCATCAGTTGGATAGTTAATTGAGTGGGGC	PLE 1 (100%) PLE 4 (93.75%) PLE 5 (100%)

	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
		2	AAGACGTGACAGCAGTGATCGACTTTATAACA	PLE 2 (100%)
ICP1_2012_A	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TGTTAGAGTCGGTAGTATCTGGATGATCGATA	PLE 2 (100%) PLE 3 (100%) PLE 4 (100%)
		3	TTTTGAAACTATTGACAGAAGGTTGGGAACCT	-
		4	TTCAAAATCTTCCGATACATAACTAGCAAGTT	PLE 3 (100%)
		5	TCACAATCAGCTATAAGCCCTGCATTTTCAAT	PLE 2 (100%)
		6	TTGTAGTGATGACATAATCTCGTCTCGACTCA	PLE 2 (100%)
		7	AATTGTCGAAGATGGTGAGGCACTAGCTACAC	PLE 1 (100%) PLE 2 (100%)
		8	TGCGCAGCCACATCACAACACACTGTAAAAAT	PLE 1 (100%) PLE 4 (96.88%) PLE 5 (93.75%)
		9	ACAAAACCTTAATAGGGACAAAAGTTATTAAA	PLE 1 (100%) PLE 3 (84.38%) PLE 4 (84.38%) PLE 5 (84.38%)
	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
		2	TTTTACGCAAAGTAGGATCGAGTGTTGCGAAC	*V. cholerae* (100%)
ICP1_2015_A	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TGTTAGAGTCGGTAGTATCTGGATGATCGATA	PLE 2 (100%) PLE 3 (100%) PLE 4 (100%)
		3	TCACAATCAGCTATAAGCCCTGCATTTTCAAT	PLE 2 (100%)
		4	TTGTAGTGATGACATAATCTCGTCTCGACTCA	PLE 2 (100%)

		5	AAGCAGAACTCACCGCCGAAGTGGAACAGCGT	PLE 1 (100%) PLE 2 (100%) PLE 3 (84.38%) PLE 5 (81.25%)
		6	TTTTGTTGTTATTTGTTATTTTGAATCAATCA	PLE 1 (100%) PLE 5 (100%)
		7	TAGTTTCACCGTGCTATTATTCTCGACAACCA	PLE 1 (100%) PLE 2 (87.5%)
	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
		2	TTTTACGCAAAGTAGGATCGAGTGTTGCGAAC	*V. cholerae* (100%)
ICP1_2016_A	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TCACAATCAGCTATAAGCCCTGCATTTTCAAT	PLE 2 (100%)
		3	TTGTAGTGATGACATAATCTCGTCTCGACTCA	PLE 2 (100%)
		4	AAGCAGAACTCACCGCCGAAGTGGAACAGCGT	PLE 1 (100%) PLE 2 (100%) PLE 3 (84.38%) PLE 5 (81.25%)
		5	TTGCATCAGTTGGATAGTTAATTGAGTGGGGC	PLE 1 (100%) PLE 4 (93.75%) PLE 5 (100%)
	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
		2	AAGACGTGACAGCAGTGATCGACTTTATAACA	PLE 2 (100%)
ICP1_2017_A	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TCACAATCAGCTATAAGCCCTGCATTTTCAAT	PLE 2 (100%)
		3	AAGCAGAACTCACCGCCGAAGTGGAACAGCGT	PLE 1 (100%) PLE 2 (100%) PLE 3 (84.38%) PLE 5 (81.25%)
		4	TTGCATCAGTTGGATAGTTAATTGAGTGGGGC	PLE 1 (100%) PLE 4 (93.75%) PLE 5 (100%)

	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
ICP1_2017_B	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TCACAATCAGCTATAAGCCCTGCATTTTCAAT	PLE 2 (100%)
		3	TTGCATCAGTTGGATAGTTAATTGAGTGGGGC	PLE 1 (100%) PLE 4 (93.75%) PLE 5 (100%)
		4	TTCACGGGTAGCAACAGGGTAATAAACCAATA	PLE 1 (100%) PLE 2 (100%) PLE 4 (93.75%) PLE 5 (100%)
	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
		2	AAGACGTGACAGCAGTGATCGACTTTATAACA	PLE 2 (100%)
ICP1_2017_C	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TGTTAGAGTCGGTAGTATCTGGATGATCGATA	PLE 2 (100%) PLE 3 (100%) PLE 4 (100%)
		3	TCACAATCAGCTATAAGCCCTGCATTTTCAAT	PLE 2 (100%)
		4	TTGTAGTGATGACATAATCTCGTCTCGACTCA	PLE 2 (100%)
		5	AAGCAGAACTCACCGCCGAAGTGGAACAGCGT	PLE 1 (100%) PLE 2 (100%) PLE 3 (84.38%) PLE 5 (81.25%)
		6	TTGCATCAGTTGGATAGTTAATTGAGTGGGGC	PLE 1 (100%) PLE 4 (93.75%) PLE 5 (100%)
	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
		2	AAGACGTGACAGCAGTGATCGACTTTATAACA	PLE 2 (100%)
ICP1_2017_D	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)

		2	TTGTAGTGATGACATAATCTCGTCTCGACTCA	PLE 2 (100%)
		3	AAGCAGAACTCACCGCCGAAGTGGAACAGCGT	PLE 1 (100%) PLE 2 (100%) PLE 3 (84.38%) PLE 5 (81.25%)
		4	TTGCATCAGTTGGATAGTTAATTGAGTGGGGC	PLE 1 (100%) PLE 4 (93.75%) PLE 5 (100%)
	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
		2	AAGACGTGACAGCAGTGATCGACTTTATAACA	PLE 2 (100%)
ICP1_2017_E	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TTGTAGTGATGACATAATCTCGTCTCGACTCA	PLE 2 (100%)
		3	AAGCAGAACTCACCGCCGAAGTGGAACAGCGT	PLE 1 (100%) PLE 2 (100%) PLE 3 (84.38%) PLE 5 (81.25%)
		4	TTGCATCAGTTGGATAGTTAATTGAGTGGGGC	PLE 1 (100%) PLE 4 (93.75%) PLE 5 (100%)
	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
		2	C	PLE 2 (100%)
ICP1_2017_F	CR 1	1	CATTGCAACTATGCAAAATGATGAAGCTAAAA	PLE 3 (100%) PLE 4 (100%) PLE 5 (100%)
		2	TCACAATCAGCTATAAGCCCTGCATTTTCAAT	PLE 2 (100%)
		3	AAGCAGAACTCACCGCCGAAGTGGAACAGCGT	PLE 1 (100%) PLE 2 (100%) PLE 3 (84.38%) PLE 5 (81.25%)
		4	TTGCATCAGTTGGATAGTTAATTGAGTGGGGC	PLE 1 (100%) PLE 4 (93.75%) PLE 5 (100%)

	CR 2	1	GTGTATTGCTTGCAGTGGGTTACACACAAGAA	PLE 2 (84.28%) PLE 3 (87.5%) PLE 4 (90.62%)
		2	AAGACGTGACAGCAGTGATCGACTTTATAACA	PLE 2 (100%)

[a] Spacers with no matches are indicated with a '-'.

Supplementary Table 3.6: Number of spacers obtained in the high-throughput spacer acquisition assay and their respective targets

	Replicate 1		Replicate 2		Replicate 3		Averages (%)	
Target Region	Unique	All spacers	Unique	All spacers	Unique	All spacers	Unique	All spacers
PLE 1	1,760	1,113,858	1,697	1,408,934	1,625	1,245,537	1,694 (36.3)	1,256,110 (96)
Chr I	424	4,194	443	6,870	392	4,534	420 (9)	5,199 (0.4)
Chr II	2,623	32,981	2,562	60,360	2,487	47,634	2,557 (54.7)	46,992 (3.6)
Total	4,807	1,151,033	4,702	1,476,164	4,504	1,297,705	4,707	1,308,340

Spacer Lengths	All spacers	All spacers	All spacers	All spacers
31bp	2,758	3,711	3,322	3,264 (0.2)
32bp	1,106,376	1,413,306	1,245,939	1,255,207 (93)
33bp	83,348	101,746	87,910	91,001 (6.7)
Total*	1,192,482	1,518,763	1,337,171	1,349,472

*Not all putative spacers mapped to the targets.

Supplementary Table 3.7: Number of additional putative engineered spacer binding sites on the *V. cholerae* small chromosome at decreasing levels of identity with the target.

Engineered spacer (kb)	100% identity	90% identity	80% identity
-93	1	0	0
-71	1	94	14
-53	1	28	72
-46	1	39	64
-22	1	0	0
-2.5	1	0	0
-1.5	1	0	0
0.5	1	1	2
22.5	1	0	0
437	1	0	0

References

1. Comeau, A. M. *et al.* Exploring the prokaryotic virosphere. *Res Microbiol* **159,** 306–313 (2008).
2. Brüssow, H., Canchaya, C. & Hardt, W.-D. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiol Mol Biol Rev* **68,** 560–602 (2004).
3. Ramisetty, B. C. M. & Sudhakari, P. A. Bacterial 'Grounded' Prophages: Hotspots for Genetic Renovation and Innovation. *Front. Genet.* **10,** 65 (2019).
4. Bobay, L.-M., Touchon, M. & Rocha, E. P. C. Pervasive domestication of defective prophages by bacteria. *Proc. Natl. Acad. Sci.* **111,** 12127–12132 (2014).
5. Penadés, J. R. & Christie, G. E. The Phage-Inducible Chromosomal Islands: A Family of Highly Evolved Molecular Parasites. *Annu. Rev. Virol.* **2,** 181–201 (2015).
6. Úbeda, C. *et al.* SaPI mutations affecting replication and transfer and enabling autonomous replication in the absence of helper phage. *Mol. Microbiol.* **67,** 493–503 (2008).
7. Tormo-Más, M. A. *et al.* Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature* **465,** 779–782 (2010).
8. Bowring, J. *et al.* Pirating conserved phage mechanisms promotes promiscuous staphylococcal pathogenicity island transfer. *Elife* **6,** e26487 (2017).
9. Mir-Sanchis, I. *et al.* Control of *Staphylococcus aureus* pathogenicity island excision. *Mol. Microbiol.* **85,** 833–845 (2012).
10. Ruzin, A., Lindsay, J. & Novick, R. P. Molecular genetics of SaPI1 - a mobile pathogenicity island in *Staphylococcus aureus*. *Mol Microbiol.* **41,** 365–377 (2001).
11. Poliakov, A. *et al.* Capsid Size Determination by *Staphylococcus aureus* Pathogenicity Island SaPI1 Involves Specific Incorporation of SaPI1 Proteins into Procapsids. *J. Mol. Biol.* **380,** 465–475 (2008).
12. Ubeda, C. *et al.* Specificity of staphylococcal phage and SaPI DNA packaging as revealed by integrase and terminase mutations. *Mol Microbiol.* **72,** 98–108 (2009).
13. Lindsay, J. a., Ruzin, A., Ross, H. F., Kurepina, N. & Novick, R. P. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol. Microbiol.* **29,** 527–543 (1998).
14. Braga, L. P. P., Soucy, S. M., Amgarten, D. E., Silva, A. M. & Setubal, J. C. Bacterial Diversification in the Light of the Interactions with Phages: The Genetic Symbionts and Their Role in Ecological Speciation. *Front. Ecol. Evol.* **6,** (2018).
15. Keen, E. C. *et al.* Novel 'Superspreader' Bacteriophages Promote Horizontal Gene Transfer by Transformation. *MBio* **8,** e02115-16 (2017).
16. Rodriguez-Valera, F. *et al.* Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* **7,** 828–36 (2009).
17. Manrique, P., Dills, M. & Young, M. J. The Human Gut Phage Community and Its Implications for Health and Disease. *Viruses* **9,** E141 (2017).
18. Faruque, S. M. *et al.* Self-limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. *Proc. Natl. Acad. Sci.* **102,** 6119–6124 (2005).
19. 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).
20. Nelson, E. J., Harris, J. B., Morris, J. G., Calderwood, S. B. & Camilli, A. Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nat. Rev. Microbiol.* **7,** 693–702 (2009).
21. Angermeyer, A., Das, M. M., Singh, D. V. & Seed, K. D. Analysis of 19 Highly Conserved *Vibrio cholerae* Bacteriophages Isolated from Environmental and Patient Sources Over a Twelve-Year Period. *Viruses* **10,** 299 (2018).
22. O'Hara, B. J., Barth, Z. K., McKitterick, A. C. & Seed, K. D. A highly specific phage defense system is a conserved feature of the *Vibrio cholerae* mobilome. *PLoS Genet.* **13,** e1006838 (2017).
23. Barth, Z. K., Angermeyer, A., Silva, T. & Seed, K. D. Replication of a viral satellite bolsters immunity against phage. *In preperation*
24. Seed, K. D., Lazinski, D. W., Calderwood, S. B. & Camilli, A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494,** 489–91 (2013).
25. Naser, I. Bin, Hoque, M. M., Nahid, M. A., Rocky, M. K. & Faruque, S. M. Analysis of the CRISPR-Cas system in bacteriophages active on epidemic strains of *Vibrio cholerae* in Bangladesh. *Sci. Rep.* **7,** 14880 (2017).
26. Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172,** 1239–1259 (2018).
27. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J Mol Evol* **60,** 174–182 (2005).
28. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).
29. McKitterick, A. C. & Seed, K. D. Anti-phage islands force their target phage to directly mediate island excision and spread. *Nat. Commun.* **9,** 2348 (2018).
30. Dy, R. L., Richter, C., Salmond, G. P. C. & Fineran, P. C. Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annu. Rev. Virol.* **1,** 307–331 (2014).
31. Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **11,** 675–87 (2013).
32. Koskella, B. & Brockhurst, M. A. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **38,** 916–931 (2014).
33. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8,** 317–327 (2010).
34. Ram, G. *et al.* Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc. Natl. Acad. Sci.* **109,** 16300–16305 (2012).
35. Ram, G., Chen, J., Ross, H. F. & Novick, R. P. Precisely modulated pathogenicity island interference with late phage gene transcription. *Proc. Natl. Acad. Sci.* **111,** 14536–41 (2014).
36. Seed, K. D. *et al.* Evolutionary consequences of intra-patient phage predation on

microbial populations. *Elife* **3,** e03497 (2014).
37. D'Herelle, F. & Malone, R. H. A Preliminary report of work carried out by the cholera bacteriophage enquiry. *Ind Med Gaz* 614–616 (1927).
38. Smith, M. C. M. Phage-encoded Serine Integrases and Other Large Serine Recombinases. 1–19 (2015). doi:10.1128/microbiolspec.MDNA3-0059-2014.f1
39. Thorpe, H. M. & Smith, M. C. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci.* **95,** 5505–5510 (1998).
40. Kuhstoss, S. & Rao, R. N. Analysis of the integration function of the Streptomycete bacteriophage φC31. *J. Mol. Biol.* **222,** 897–908 (1991).
41. Bibb, L. A., Hancox, M. I. & Hatfull, G. F. Integration and excision by the large serine recombinase ϕRv1 integrase. *Mol. Microbiol.* **55,** 1896–1910 (2005).
42. Khaleel, T., Younger, E., Mcewan, A. R., Varghese, A. S. & Smith, M. C. M. A phage protein that binds ϕC31 integrase to switch its directionality. *Mol. Microbiol.* **80,** 1450–1463 (2011).
43. Savinov, A., Pan, J., Ghosh, P. & Hatfull, G. F. The Bxb1 gp47 recombination directionality factor is required not only for prophage excision, but also for phage DNA replication. *Gene* **495,** 42–48 (2012).
44. Fogg, P. C. M., Colloms, S., Rosser, S., Stark, M. & Smith, M. C. M. New applications for phage integrases. *J. Mol. Biol.* **426,** 2703–2716 (2014).
45. Ghosh, P., Wasil, L. R. & Hatfull, G. F. Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biol.* **4,** 0964–0974 (2006).
46. Frígols, B. *et al.* Virus Satellites Drive Viral Evolution and Ecology. *PLoS Genet.* **11,** 3–5 (2015).
47. Misiura, A. *et al.* Roles of two large serine recombinases in mobilizing the methicillin-resistance cassette SCC*mec*. *Mol. Microbiol.* **88,** 1218–1229 (2013).
48. Yen, M., Cairns, L. S. & Camilli, A. A cocktail of three virulent bacteriophages prevents *Vibrio cholerae* infection in animal models. *Nat. Commun.* **8,** 14187 (2017).
49. *Bacteriophages : methods and protocols*. *Methods in molecular biology* (Humanan Press, 2009).
50. Dalia, A. B., Lazinski, D. W. & Camilli, A. Identification of a Membrane-Bound Transcriptional Regulator That Links Chitin and Natural Competence in *Vibrio cholerae*. *MBio* **5,** e01028-13 (2014).
51. Fürste, J. P. *et al.* Molecular cloning of the plasmid RP4 primase region in a multi-host-range tacP expression vector. *Gene* **48,** 119–131 (1986).
52. Topp, S. *et al.* Synthetic riboswitches that induce gene expression in diverse bacterial species. *Appl. Environ. Microbiol.* **76,** 7881–7884 (2010).
53. Box, A. M., McGuffie, M. J., O'Hara, B. J. & Seed, K. D. Functional analysis of bacteriophage immunity through a Type I-E CRISPR-Cas system in *Vibrio cholerae* and its application in bacteriophage genome engineering. *J. Bacteriol.* **198,** 578–90 (2016).
54. Ouellette, S. P., Karimova, G., Davi, M. & Ladant, D. Analysis of Membrane Protein Interactions with a Bacterial Adenylate Cyclase–Based Two-Hybrid (BACTH) Technique. 1–24 (2017).
55. Marchler-Bauer, A. *et al.* CDD/SPARCLE: Functional classification of proteins via

subfamily domain architectures. *Nucleic Acids Res.* **45,** D200–D203 (2017).
56. Alva, V., Nam, S.-Z., Söding, J. & Lupas, A. N. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* **44,** W410–W415 (2016).
57. Crooks, G., Hon, G., Chandonia, J. & Brenner, S. WebLogo: a sequence logo generator. *Genome Res* **14,** 1188–1190 (2004).
58. Bawono, P. & Heringa, J. PRALINE: A Versatile Multiple Sequence Alignment Toolkit. *Methods Mol Bio* **1079,** 245–62 (2014).
59. Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43,** W580–W584 (2015).
60. Koonin, E. V & Krupovic, M. Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat Rev Genet* **16,** 184–192 (2015).
61. Bondy-Denomy, J. & Davidson, A. R. When a virus is not a parasite: The beneficial effects of prophages on bacterial fitness. *J. Microbiol.* **52,** 235–242 (2014).
62. Pires, D. P., Cleto, S., Sillankorva, S., Azeredo, J. & Lu, T. K. Genetically Engineered Phages: a Review of Advances over the Last Decade. *Microbiol. Mol. Biol. Rev.* **80,** 523–543 (2016).
63. Hinton, D. M. *et al.* Transcriptional takeover by σ appropriation: Remodelling of the σ70 subunit of Escherichia coli RNA polymerase by the bacteriophage T4 activator MotA and co-activator AsiA. *Microbiology* **151,** 1729–1740 (2005).
64. Warner, H. R., Snustad, D. P., Jorgensen, S. E. & Koerner, J. F. Isolation of Bacteriophage T4 Mutants Defective in the Ability to Degrade Host Deoxyribonucleic Acid. *J Virol* **5,** 700–708 (1970).
65. Hercules, K., Munro, J. L., Mendelsohn, S. & Wiberg, J. S. Mutants in a Nonessential Gene of Bacteriophage T4 Which Are Defective in the Degradation of Escherichia coli Deoxyribonucleic Acid. *J Virol* **7,** 95–105 (1971).
66. Penadés, J. R., Chen, J., Quiles-Puchalt, N., Carpena, N. & Novick, R. P. Bacteriophage-mediated spread of bacterial virulence genes. *Curr. Opin. Microbiol.* **23,** 171–8 (2015).
67. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. *J. Bacteriol.* **193,** 6039–6056 (2011).
68. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science (80-.).* **359,** eaar4120 (2018).
69. McKitterick, A. C., LeGault, K., Angermeyer, A., Alam, M. & Seed, K. D. Competition between mobile genetic elements drives optimization of a phage-encoded CRISPR-Cas system: Insights from a natural arms-race. *Phil. Trans. R. Soc. B* **374,** 20180089 (2019).
70. Saikrishnan, K., Powell, B., Cook, N. J., Webb, M. R. & Wigley, D. B. Mechanistic Basis of 5′-3′ Translocation in SF1B Helicases. *Cell* **137,** 849–859 (2009).
71. Byrd, A. K. & Raney, K. D. Structure and function of Pif1 helicase. *Biochem. Soc. Trans.* **45,** 1159–1171 (2017).
72. Singleton, M. R., Dillingham, M. S., Gaudier, M., Kowalczykowski, S. C. & Wigley, D. B. Crystal structure of RecBCD enzyme reveals a machine for processing DNA

breaks. *Nature* **432,** 187–193 (2004).
73. He, X. *et al.* The T4 phage SF1B helicase Dda is structurally optimized to perform DNA strand separation. *Structure* **20,** 1189–1200 (2012).
74. Byrd, A. K. & Raney, K. D. Displacement of a DNA binding protein by Dda helicase. *Nucleic Acids Res.* **34,** 3020–3029 (2006).
75. Brister, J. R. Origin Activation Requires both Replicative and Accessory Helicases during T4 Infection. *J. Mol. Biol.* **377,** 1304–1313 (2008).
76. Vector Bangladesh Map. Available at: https://www.vecteezy.com/vector-art/107097-vector-bangladesh-map.
77. Blair, L. P., Tackett, A. J. & Raney, K. D. Development and evaluation of a structural model for SF1B helicase dda. *Biochemistry* **48,** 2321–2329 (2009).
78. Kulczyk, A. W. & Richardson, C. C. The Replication System of Bacteriophage T7. *Enzymes* **39,** 89–136 (2016).
79. Crow, M. S., Lum, K. K., Sheng, X., Song, B. & Cristea, I. M. Diverse mechanisms evolved by DNA viruses to inhibit early host defenses. *Crit. Rev. Biochem. Mol. Biol.* **51,** 452–481 (2016).
80. Rivas, H. G., Schmaling, S. K. & Gaglia, M. M. Shutoff of host gene expression in influenza A virus and herpesviruses: Similar mechanisms and common themes. *Viruses* **8,** 102 (2016).
81. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).
82. Deatherage, D. E. & Barrick, J. *Evolved Microbes from Next-Generation Sequencing Data Using breseq, in Engineering and Analyzing Multicellular Systems: Methods and Protocols*. (Springer, 2014).
83. Edgar, R. C. MUSCLE : multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32,** 1792–1797 (2004).
84. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948 (2007).
85. Guindon, S., Lethiec, F., Duroux, P. & Gascuel, O. PHYML Online — a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* **33,** W557–W559 (2005).
86. Parikka, K. J., Le Romancer, M., Wauters, N. & Jacquet, S. Deciphering the virus-to-prokaryote ratio (VPR): Insights into virus-host relationships in a variety of ecosystems. *Biol. Rev.* **92,** 1081–1100 (2017).
87. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315,** 1709–1712 (2007).
88. Sinkunas, T. *et al.* Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **30,** 1335–1342 (2011).
89. Koonin, E. V. & Makarova, K. S. Mobile genetic elements and evolution of CRISPR-Cas systems: all the way there and back. *Genome Biol. Evol.* **9,** 2812–2825 (2017).
90. Shmakov, Sergey; Sitnik, Vassilii; Makarova, Kira; Wolf, Yuri; Severinov, Konstantin; Koonin, E. The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes. *MBio* **8,** e01397-17 (2017).
91. Palmer, K. L. & Gilmore, M. S. Multidrug-Resistant Enterococci Lack CRISPR-Cas. *MBio* **1,** e00227-10 (2010).

92. Bikard, D., Hatoum-Aslan, A., Mucida, D. & Marraffini, L. A. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* **12,** 177–186 (2012).
93. Novick, R. P. & Ram, G. The Floating (Pathogenicity) Island: A Genomic Dessert. *Trends Genet.* **32,** 114–126 (2016).
94. Haurzitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science* **329,** 1355–1358 (2010).
95. Richter, C. & Fineran, P. C. The subtype I-F CRISPR–Cas system influences pathogenicity island retention in *Pectobacterium atrosepticum* via crRNA generation and Csy complex formation. *Biochem. Soc. Trans.* **41,** 1468–1474 (2013).
96. Fagerlund, R. D. *et al.* Spacer capture and integration by a type I-F Cas1–Cas2-3 CRISPR adaptation complex. *Proc. Natl. Acad. Sci.* **114,** E5122–E5128 (2017).
97. Rollins, M. F. *et al.* Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. *Proc. Natl. Acad. Sci.* **114,** E5113–E5121 (2017).
98. Bellas, C. M., Anesio, A. M. & Barker, G. Analysis of virus genomes from glacial environments reveals novel virus groups with unusual host interactions. *Front. Microbiol.* **6,** (2015).
99. Chénard, C., Wirth, J. F. & Suttle, C. A. Viruses Infecting a Freshwater Filamentous Cyanobacterium (*Nostoc* sp.) Encode a Functional CRISPR Array and a Proteobacterial DNA Polymerase B. *MBio* **7,** e00667-16 (2016).
100. Barker, A., Clark, C. A. & Manning, P. A. Identification of VCR, a repeated sequence associated with a locus encoding a hemagglutinin in *Vibrio cholerae* O1. *J. Bacteriol.* **176,** 5450–5458 (1994).
101. Staals, R. H. J. *et al.* Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat. Commun.* **7,** 12853 (2016).
102. Vorontsova, D. *et al.* Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery. *Nucleic Acids Res.* **43,** 10848–10860 (2015).
103. Van Erp, P. B. G. *et al.* Mechanism of CRISPR-RNA guided recognition of DNA targets in *Escherichia coli*. *Nucleic Acids Res.* **43,** 8381–8391 (2015).
104. Van Houte, S. *et al.* The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532,** 385–388 (2016).
105. Richter, C. *et al.* Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.* **42,** 8516–8526 (2014).
106. Xue, C. *et al.* CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.* **43,** 10831–10847 (2015).
107. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–740 (2009).
108. Redding, S. *et al.* Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System. *Cell* **163,** 854–865 (2015).
109. Brown, M. W. *et al.* Assembly and translocation of a CRISPR-Cas primed

acquisition complex. *Cell* **175,** 934–946 (2018).
110. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).
111. Jinek, M. *et al.* A Programmable Dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
112. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13,** 722–736 (2015).
113. Yosef, I., Manor, M., Kiro, R. & Qimron, U. Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. *Proc. Natl. Acad. Sci.* **112,** 7267–72 (2015).
114. Bikard, D. *et al.* Exploiting CRISPR-cas nucleases to produce sequence-specific antimicrobials. *Nat. Biotechnol.* **32,** 1146–1150 (2014).
115. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27,** 722–736 (2017).
116. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One* **9,** e112963 (2014).
117. Lazinski, D. W. & Camilli, A. Homopolymer tail-mediated ligation PCR: A streamlined and highly efficient method for DNA cloning and library construction. *Biotechniques* **54,** 25–34 (2013).
118. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
119. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25,** 1422–1423 (2009).
120. Hu, C.-C., Hsu, Y.-H. & Lin, N.-S. Satellite RNAs and Satellite Viruses of Plants. *Viruses* **1,** 1325–1350 (2009).
121. Levine, M. M. *et al.* The Pathogenicity of Nonenterotoxigenic *Vibrio cholerae* Serogroup O1 Biotype El Tor Isolated from Sewage Water in Brazil. *J. Infect. Dis.* **145,** 296–299 (1982).
122. Gong, Z. *et al.* Isolation and Complete Genome Sequence of a Novel *Pseudoalteromonas* Phage PH357 from the Yangtze River Estuary. *Curr. Microbiol.* **74,** 832–839 (2017).
123. Luhtanen, A.-M. *et al.* Isolation and characterization of phage–host systems from the Baltic Sea ice. *Extremophiles* **18,** 121–130 (2014).
124. Petrov, V. M., Ratnayaka, S., Nolan, J. M., Miller, E. S. & Karam, J. D. Genomes of the T4-related bacteriophages as windows on microbial genome evolution. *Virol J* **7,** (2010).
125. Kauffman, K. M. *et al.* Viruses of the Nahant Collection, characterization of 251 marine Vibrionaceae viruses. *Sci. Data* **5,** 180114 (2018).
126. Ramphul, C. *et al.* Genome analysis of three novel lytic *Vibrio coralliilyticus* phages isolated from seawater, Okinawa, Japan. *Mar. Genomics* **35,** 69–75 (2017).
127. McDonough, E., Lazinski, D. W. & Camilli, A. Identification of *in vivo* regulators of the *Vibrio cholerae xds* gene using a high-throughput genetic selection. *Mol. Microbiol.* **92,** 302–315 (2014).