The antiviral activities of a bacteriophage satellite are mechanistically tied to lateral propagation

By

Zachary K Barth

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 Britt Glaunsinger
Professor Russell Vance
Professor Britt Koskella

Summer 2020

Abstract

The antiviral activities of a bacteriophage satellite are mechanistically tied to lateral propagation

By Zachary K Barth

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Kimberley Seed, Chair

Parasitism is a common evolutionary strategy present in many branches of life, and the study of parasitism has contributed substantially to evolutionary theory. Bacteriophages, viruses that infect bacteria, are one group of parasites that have been particularly valuable models of study. Study of bacteriophages, or phages as they are also known, has vastly improved our understanding of molecular biology. The study of bacterial systems that defend against phages has provided incredibly useful tools for genetic research such as restriction enzymes and CRISPR-Cas. A particularly interesting form of anti-phage defense is hyperparasitism, parasitism of a parasite, by bacteriophage satellites that are endogenous to the bacterial genome. Bacteriophage satellites rely on bacteriophage infection for horizontal mobility while reducing or restricting production of progeny bacteriophage. This allows them to function as immune systems for the cellular populations that encode them. Within this thesis, we identify mechanisms of parasitism and bacteriophage interference by the phage inducible chromosomal island-like element or PLE, a bacteriophage satellite found within some strains of *Vibrio cholerae*. Using deep sequencing approaches, we have outlined DNA replication and transcriptional programs for both the PLE, and the phage it parasitizes, ICP1. Molecular and phenotypic approaches were also applied to establish how PLE parasitizes ICP1 for its own genome replication and what effect this has on the life cycle of ICP1. We have found that PLE requires ICP1 not just for induction of gene expression, but also for replication of the PLE genome. Consistent with relying on ICP1 machinery for replication and mobilization, PLE does not broadly interfere with ICP1 transcription. The sole exception to this is repression of ICP1 capsid morphogenesis genes, consistent with the PLE remodeling ICP1 virions to fit its smaller genomes. The experiments presented here, along with other recent data, suggest that PLE inhibits ICP1 through multiple mechanisms which have dual functions of also boosting PLE reproduction and mobilization. This thesis provides a foundation for understanding the parasitism strategies of ICP1 and PLE, as well as the mechanisms through which PLE restricts ICP1. Our results suggest that PLE has adapted to interfere with the replication program of ICP1 only to an extent that is beneficial to the PLE, and fit within a context pattern of pathogen and parasite evolutionary patterns.

Table of Contents

Chapter 1

An introduction to parasitism, parasitoids, and bacteriophages

The Vermin only teaze and pinch
Their Foes superior by an Inch.
So, Nat'ralists observe, a Flea
Hath smaller Fleas that on him prey,
And these have smaller yet to bite 'em,
And so proceed *ad infinitum*:
Thus ev'ry Poet, in his Kind
Is bit by him that comes behind
-Jonathan Swift

These lines appear as part of Jonathan Swift's satirical poem "On Poetry: A Rhapsody." Whether they provide an apt metaphor for a poet's creative struggles is beyond the purview of science; however, they are a serviceable introduction to parasitism and antagonistic coevolution. Like the successive poets in the last couplet, evolutionary foes, be they prey and predator or host and parasite, are engaged in a cycle of adaptation and counter adaptation. In the case of the most intimate symbionts, viruses and their host cells, there can be and often is literal theft of innovation in the form of horizontal gene transfer [1,2]: because host genes have already evolved to interface with the host system, they can provide a shortcut to subverting those systems when acquired and repurposed by pathogens. Stepping back from the dynamics of coevolution, these lines also speak to the ubiquity of parasitism. Parasitism is a tremendously successful life strategy to the point that parasites often have their own parasites.

What we understand to be fleas today are not actually known to be parasitized by other fleas, but parasitic relationships between insects are relatively common. In choosing the 'flea' as the subject of his poem, Jonathan Swift may have been influenced by contemporary entomologists. Among those entomologists was Antonio Vallisneri. Describing small black 'flies' emerging from dead 'fleas', Vallisneri wrote:

> I found that that Little Fly had very kindly laid an egg under the belly of the other insect, and was doing the same thing with the others. Having this piece of good news, it was not hard for me to find out why Little Flies flew out of those Fleas which looked dead, in view of the fact that lots of little maggots hatch out of these eggs and, as soon as they hatch, the [sic] pierce the bellies of the Fleas underneath which they have been laid, and entering their bodies, they use them both as food and as a room: and when they are well enough nourished and grown, they make a chrysalis inside the same Flea, then in less than a Month the Little Flies coming flying out, one from each Flea. [3]

Today, Vallisneri's 'Flea' is recognized to be an aphid, a parasite of plants, and the 'Little Flies' to be parasitoid wasps [3]. Parasitoid wasps lay eggs on or in the eggs, pupae, or bodies of their prey. The eggs give rise to larvae which feed and develop inside their hosts. In some cases, a single wasp emerges, in others a single egg gives rise to many wasps that erupt from their host at the same time. Some wasps are

generalists that can feed on many species while others have evolved to use only one host [4]. The use of the descriptor ‘parasitoid’ rather than ‘parasitic’ for these wasps is meant to convey that their life cycle is similar to but distinct from parasitism as its usually understood. Like parasites, parasitoids live in close association with their host; however, like a predator eating its prey, the parasitoid always kills its host. While parasites often contribute to disease and eventual death in their hosts, host death is generally a dead end for the infection or infestation. For parasitoids, host death, at the correct moment, is part of the parasitoid’s life cycle.

Parasitoid wasps have been an important subject in biology stretching back to the beginnings of modern science. The observation that a wasp’s sting of a caterpillar preceded the caterpillar’s infestation by new wasps was used as early evidence against spontaneous generation [5]. The perceived cruelty of the parasitoid wasp’s life cycle helped convince Charles Darwin that species were not created as they currently exist by a benevolent deity [6]. Darwin evokes parasitic wasps multiple times in *On the Origins of Species*, and cites that while many wasps are strictly parasitic, some hunting wasps are facultative parasites of other wasp nests, suggesting that generalists could evolve into specialists [7].

Parasitic lifestyles are not just useful for understanding evolution; they are also evolutionarily successful. Speaking to the success of parasites and parasitoids, recent work suggests that parasitoid wasps are the most speciose group of insects, which in turn are the most speciose group of animals [8]. Like Swift’s imaginary fleas, hyperparasitism, a biological relationship in which parasites host their own parasites, is common among these wasps. Gall wasps are wasps that parasitize plants by inducing the formation of galls, chambered tumors in which larvae can feed and grow. In oak galls, up to six levels of parasitism have been seen with the gall wasp feeding on the oak tree providing the first level, and the other five levels occupied by other wasps [9]. While this is not quite proceeding *ad infinitum*, it should be noted that potential higher levels of parasitism such as that by microscopic wasps too small to see with the naked eye, fungi, nematodes, viruses, and other microscopic parasites are not accounted for here.

Changing views from the most abundant group of animals to the most abundant group of cellular life, bacteria, provides further evidence for the success of parasitism. Humans are infected by several pathogenic bacterial species [10,11]. While the number of human pathogens is miniscule in light of the full diversity of bacteria, humanity itself comprises a miniscule branch of the tree of life. It’s true that many pathogens are opportunists that most often exist as benign commensals in a tenuous détente with their hosts [12], but the existence of pathogenic specialists such as *Treponema pallidum*, and *Mycobacterium tuberculosis* suggests that there are specific bacterial parasites for different host species. While some bacterial pathogens spread repeatedly from chronically infected hosts, others have a more parisitoidal life cycle. One such example is found in toxigenic *Vibrio cholerae*, a pathogen endemic to estuarine waters and a major source of mortality in developing countries. If consumed, *V. cholerae* will colonize

the small intestine and replicate to high numbers [13]. The *V. cholerae* progeny cells induce profuse watery diarrhea to achieve mass egress from the host.

Unsurprisingly, bacteria have their own parasites and parasitoids, which exist primarily in the form of bacteriophages, viruses that infect bacteria. Bacteriophages, often referred to as phages, are also superficially similar to parasitoid wasps. Like a wasp injecting its egg into a host, phages inject their genomes into bacterial cells. The genome will then replicate, new phages are constructed inside the cell, and finally the cell will lyse and progeny phage are released, like wasps emerging from the empty shell of their host. Similar to how parasitoid wasps dominate insects in terms of number of species and number of individuals, bacteriophages outnumber bacteria cells ten to one [14], making bacteriophages the most abundant biological entities on earth.

Paralleling how observations of parasitoid wasps helped revolutionize early modern biology, phages have played a critical role in more recent biological discoveries. Experiments showing that DNA and not protein is the genetic material [15], that genes exist as regions of DNA [16,17] and those regions are read in a triplet basepair code [18] were all performed using bacteriophages.

Among the most striking aspects of parasites and parasitoids are the fantastic and gruesome changes they induce in their hosts. In some cases, wasps emerge still in the larval stage, and the host is able to survive its partial consumption and the larvae's exit of its body. The larvae begin to pupate nearby, and the host, still under the influence of infestation, becomes a guard against parasitoid species that would infest the pupating wasps [19,20] . Behavioral manipulation is not limited to wasps. Semiaquatic horsehair worms will drive their insect hosts to drown themselves [21], infection by *Toxoplasma gondii*, causes rats to no longer fear the scent of cat urine [22], and the fungus *Massospora cicadina* induces hypersexual mating behavior in male cicadas after it has replaced their genitalia with a spore discharging fungal growth [23].

Single cells lack nervous systems to manipulate, but phages and other viruses mount a similar takeover of their hosts. In order for infections to be successful, viruses must typically take over at least some of their host's gene expression, and reroute the cellular metabolism for viral production. As would be expected for this antagonistic relationship where the phage is benefitting to the detriment of the bacterial cell, bacteria have evolved defenses to thwart infection, and phages have in turn evolved counter adaptations to those defenses. Studying bacterial defenses has proven invaluable to molecular biology, and has twice provided tools that revolutionized genetic engineering, the first being restriction enzyme cloning in the 1970s [24] and the second being CRISPR-cas gene editing in the last decades [25].

A particularly striking form of antiviral defense is the cooption of viral satellites to defend cell populations. Viral satellites are hyperparasites of viruses, and as such they must parasitize viruses to move horizontally. Viral satellites have been observed in all domains of life, but it is primarily in bacteria and unicellular eukaryotes where satellites have been seen to provide partial or complete immunity against the infecting virus by

restricting the release of infectious viral particles from infected cells. While the individual infected cells die, the larger population of cells bearing the satellite survives if viruses are not released. In place of infectious viral particles being released upon lysis, virions bearing the satellite's genome are produced, which can then mobilize to other host strains.

This thesis is dedicated to understanding the parasitic relationship of a satellite and phage pair that exist in toxigenic *V. cholerae*. The phage, ICP1, has been isolated from both environmental sources and cholera patient stool samples [26]. ICP1 phages use the *V. cholerae* O1 antigen as the viral receptor for cell entry [27]. Since the O1 antigen is needed for *V. cholerae's* colonization of the human intestinal tract, its use as ICP1's viral receptor suggests that ICP1 may be adapted for propagation inside the human gut. Supporting this hypothesis, is the presence of ICP1 in cholera patient stool samples, as well as the observation that ICP1 replicates poorly in conditions mimicking the aquatic environment [28].

The satellite that preys on ICP1 is named the phage inducible chromosomal island-like element or PLE for short. Like other bacteriophage satellites, PLE excises from the bacterial chromosome following infection by its host phage, ICP1, replicates to high copy, and can be transduced to new cells following infection [29]. An in-depth summary of the PLE and its life cycle appears in the introduction of chapter 2 of this thesis. An important way that PLE differs from other bacteriophage satellites is that it completely blocks phage production when it is unimpaired by phage counter defenses. This particularly stringent restriction of ICP1 by PLE suggests that PLE is more properly defined as a parasitoid rather than a parasite of ICP1. Remarkably, this restriction also makes PLE an effective immune system for PLE encoding *V. cholerae* populations revealing a natural example of biocontrol of a pathogen (ICP1) by its host (*V. cholerae*).

The core of this thesis is two studies that look at DNA replication and gene expression in ICP1 and PLE. For both ICP1 and PLE, new insights are gleaned into the nature of their parasitic lifestyles and how they have adapted to exploit their respective hosts. While ICP1 causes dramatic changes to host cell gene expression, PLE appears adapted to make use of ICP1's reproduction program as it already exists, only causing small and targeted gene expression changes. The PLE's own gene expression does, however, appear to be sufficient to redirect DNA replication to the PLE's own replicon at the expense of ICP1 late in infection. Notably, the studies in this thesis also provide the first insights into how PLE is able to restrict ICP1. PLE appears to have several anti-ICP1 activities that act in combination to prevent ICP1 virion production. As has been seen in other phage satellite systems, these mechanisms appear to block packaging of full phage genomes. This work provides a foundation for understanding PLEs parasitism of ICP1, reveals key aspects of ICP1's life cycle, and brings us closer to understanding this remarkable system where a human pathogen is preyed on by a virus that can be thwarted by its own parasite.

Chapter 2

Genome replication dynamics of a bacteriophage and its satellite reveal strategies for parasitism and viral restriction [30]

Zachary K Barth[1], Tania V Silvas[1], Angus Angermeyer[1], Kimberley D Seed[1,2,*]

Summary

Phage-inducible chromosomal island-like elements (PLEs) are bacteriophage satellites found in *Vibrio cholerae*. PLEs parasitize the lytic phage ICP1, excising from the bacterial chromosome, replicating, and mobilizing to new host cells following cell lysis. PLEs protect their host cell populations by completely restricting the production of ICP1 progeny. Previously, it was found that ICP1 replication was reduced during PLE(+) infection. Despite robust replication of the PLE genome, relatively few transducing units are produced. We investigated if PLE DNA replication itself is antagonistic to ICP1 replication. Here we identify key constituents of PLE replication and assess their role in interference of ICP1. PLE encodes a RepA_N initiation factor that is sufficient to drive replication from the PLE origin of replication during ICP1 infection. In contrast to previously characterized bacteriophage satellites, expression of the PLE initiation factor was not sufficient for PLE replication in the absence of phage. Replication of PLE was necessary for interference of ICP1 DNA replication, but replication of a minimalized PLE replicon was not sufficient for ICP1 DNA replication interference. Despite restoration of ICP1 DNA replication, non-replicating PLE remained broadly inhibitory against ICP1. These results suggest that PLE DNA replication is one of multiple mechanisms contributing to ICP1 restriction.

INTRODUCTION

Viral satellites are found in all domains of life and can have a profound impact on their helper viruses and their host cells [31–33]. These sub-viral agents are known to worsen disease in humans [34] as well as plants [35], provide bacterial pathogens with toxins necessary for virulence [36], and serve as anti-viral immune systems in both single celled eukaryotes [37] and bacteria [38]. As the parasites of viruses, satellites face distinct challenges in their life cycles. Viruses typically need to subvert host cell nucleic acid metabolism in order to replicate their genome. In turn, viral satellites must find a way to subvert the subverters, so that the satellite's genome can be replicated and mobilized in addition to, or to the exclusion of, the helper virus.

Within bacteria, four phylogenetically unrelated families of tailed-bacteriophage satellites have been discovered. These include satellite phage P4 and its relatives found in *Escherichia coli* [39,40], the phage inducible chromosomal islands (PICIs) widespread throughout Firmicutes [41], the PICI-like elements (PLEs) found in epidemic isolates of *V. cholerae* [29], and the recently discovered Gram-negative PICIs found in Enterobacterialles and Pasturellales [42]. Certain details of the life cycles of PLEs and their helper phage, ICP1, distinguish PLEs from other bacteriophage satellites. Both P4 and the well characterized subfamily of PICIs referred to as staphylococcal pathogenicity islands (SaPIs) confer partial restriction of their helper phages [43,44]. In contrast, PLEs completely restrict ICP1 production when they are able to progress through their replication cycle [29]. This allows PLEs to function as effective abortive infection systems: individual ICP1 infected cells die, but since no phage are produced, the population as a whole is protected [29]. PLEs' more severe restriction of their helper phage likely relates to ICP1's life cycle. ICP1 is only known to produce lytic infections that kill the host cell [26]. In contrast, both P4 and PICIs parasitize temperate phages which occasionally integrate into the genomes of the cells they infect. For P4 and PICIs, it is not uncommon to find a helper phage and its satellite lysogenizing the same strain [39,45]. Since satellites rely on their helpers for mobilization, there can be intrinsic benefits to a low level of helper phage production that allows for co-lysogeny. If ICP1 kills every cell it can infect, cells that are potential hosts for PLEs, then it is to the PLEs' benefit to completely restrict the production of infectious ICP1 progeny.

PLEs' use of ICP1 as a helper virus also has implications for PLEs' genome replication strategy. P4's helper phage is known to rely on host encoded machinery [46], and while the replication of PICI helper phages have not been extensively characterized, comparative genomics suggest that some of the characterized PICI helpers hijack host cell replication machinery [47,48]. Similar to their helpers, both P4 and PICIs must redirect cellular machinery to the satellite genomes [49–51]. The better characterized PICIs (i.e. SaPIs) have replication initiators that possess helicase activity and primases. P4 makes use of the same activities for its own replication initiation, but the initiating helicase and the primase are fused into a single protein. When these replication genes are expressed they are sufficient to drive autonomous satellite replication within the host cell [49,51] Like many well-studied lytic phages, ICP1 differs from the helper phages exploited by P4 and PICIs by encoding its own replication machinery [26]. PLEs must

therefore use a separate DNA polymerase from ICP1, or hijack ICP1's DNA polymerase for their own replication. Either possibility provides a novel twist to bacteriophage satellite DNA replication.

For its own replication, ICP1 encodes a Pol-I type DNA polymerase and a helicase-primase with a GP4d helicase domain like the *E. coli* phage T7[52–54]. The T7 replisome is one of the best characterized replisomes and is simpler in its components than most other replication complexes. Only four proteins are needed to reconstitute T7's replisome *in vitro* [53]: a DNA polymerase; host encoded thioredoxin, which acts as a processivity factor for the polymerase; a helicase-primase, which in addition to possessing both helicase and primase activity has single stranded DNA binding activity and loads the DNA polymerase; and a single stranded DNA binding protein that aids in replisome assembly and is necessary for lagging strand synthesis. The relative simplicity of the ICP1 replisome may make it an attractive target for exploitation by PLE. Indeed, PLE replication through use of ICP1's replisome would then be in line with PLE's reliance on ICP1 for multiple steps in the PLE life cycle. To excise PLE from the bacterial chromosome, the PLE integrase requires an ICP1-encoded recombination directionality factor (rdf) [55], and PLE also requires the same viral receptor as ICP1 for transduction [29], suggesting that PLE is packaged into ICP1 capsids just as P4 and PICIs are packaged into the capsids of their helper phage [32].

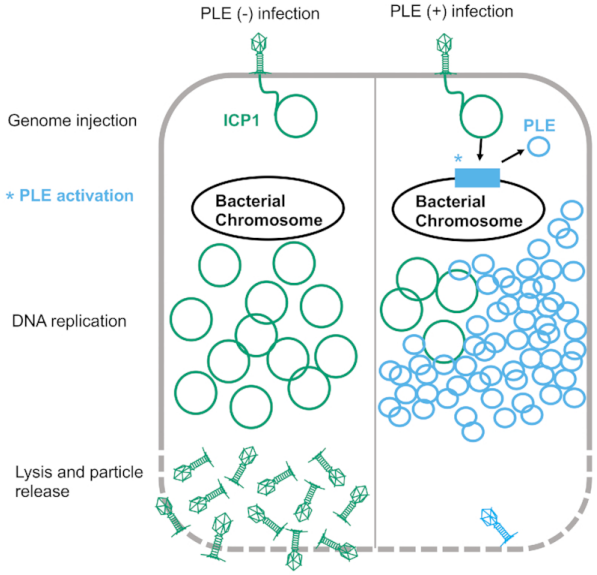


Figure 1.1. Model of ICP1 infection in PLE(−) and PLE(+) V. cholerae. ICP1 injects its DNA into V. cholerae; prior to DNA replication, ICP1 activity leads to PLE activation and excision. ICP1 DNA replication is reduced in the PLE(+) cell where the PLE replicates to high copy. Finally, the cell lyses and releases infectious particles. No ICP1 particles and a low number of PLE transducing particles are released from the PLE(+) cell.

PLE's severe parasitism of ICP1 has necessitated ICP1's evolution of counter defenses. ICP1 host range on different PLEs varies among ICP1 isolates in a manner reminiscent of host-parasite co-evolution [29]. So far, five distinct PLEs have been identified, and there has been a temporal succession of these elements, with a new PLE emerging around the same time as the previous PLE disappears from sequenced isolates. PLEs are prevalent, occurring in ~25% of *V. cholerae* isolates spanning a 60 year collection period [29]. PLE(+) *V. cholerae* have been isolated from cholera patient stool samples alongside ICP1, suggesting that ICP1 infection, and PLE parasitism of ICP1, takes place within human hosts [29,38,55,56]. ICP1 isolates appear to have multiple strategies to overcome PLE [29], but the only mechanism identified so far is a phage-encoded CRISPR-Cas system [38,56]. Among PLE genes, only the PLE integrase has

a recognized function [55], and the precise mechanism(s) by which PLEs restrict ICP1 continue to elude complete characterization.

Given the crucial role of genome replication in viral propagation, the interface of PLE and ICP1 DNA replication is likely tied to PLEs' ability to restrict ICP1. Previous work showed that PLEs can replicate upwards of 1000-fold following ICP1 infection [29] (Fig 1.1). This increase in PLE copy is accompanied by a 3 to 4-fold inhibition of ICP1 DNA replication. Curiously, PLEs do not transduce well under laboratory conditions, producing fewer than one PLE transducing unit per infected cell. Further, in these laboratory conditions, four of the five PLEs integrate seemingly randomly into one of *V. cholerae's* many *Vibrio cholerae*-repeats (VCRs), but for PLE(+) *V. cholerae* isolates from nature, each of the four PLEs always occupies the same VCR, indicating that horizontal transmission may be rare [29]. This suggests that transduction may play a minor role in the PLE life cycle, and/or that it may be infrequent. The discrepancy between robust PLE replication and poor PLE mobilization led us to investigate the requirements for PLE replication, and whether PLE may bolster its anti-phage activity through increasing its copy number.

Here, we define the replicon of PLE 1 (hereafter referred to as PLE), identifying an origin of replication and a PLE encoded replication initiation factor. The PLE replication initiator belongs to the RepA_N family of proteins, and to our knowledge is the first RepA_N protein functionally characterized in a Gram-negative bacterium. While PLE replication is not necessary to provide anti-phage immunity against ICP1, loss of PLE replication does restore the level of ICP1 DNA replication, and allows for a low level of ICP1 virion production, suggesting that PLE replication may be one of perhaps several parallel mechanisms that work in tandem to restrict ICP1.

MATERIAL AND METHODS

Strains and culture conditions

V. cholerae strains used in this study are derived from E7946. Bacteria were routinely grown on LB agar plates and in LB broth with aeration at 37°C. Antibiotics were supplemented as appropriate at the following concentrations: 75 μg/ml kanamycin, 100 μg/ml spectinomycin, 1.25 or 2.5 μg/ml chloramphenicol (*V. cholerae* for broth or plate conditions, respectively), 25 μg/ml chloramphenicol (*E. coli*), 100 μg/ml streptomycin. A detailed list of all strains used throughout this study can be found in Supplementary Table S1.1. ICP1_2006_E engineered to lack CRISPR-Cas (ΔCRISPR, Δcas2-3) [55] was used for all experiments. Phage titers were determined using a soft agar overlay method wherein ICP1 was allowed to adsorb to *V. cholerae* for 10 minutes at room temperature before the mixture was added to molten LB soft Agar (0.3%) and poured onto 100mm x 15mm LB Agar plates. Plaques were counted after overnight incubation at 37°C. Efficiency of plaquing on mutant PLE strains was determined by dividing the phage titer obtained on the mutant PLE(+) strain by the phage titer obtained on PLE(-) strain.

Generation of mutant strains and constructs

V. cholerae mutants were generated through natural transformation or *sacB* counter selection. Natural transformation was performed as described previously [57]. For gene knockouts, splicing by overlap extension (SOE) PCR was used to generate deletion constructs with a spectinomycin resistance cassette flanked by frt recombination sites. Following selection of spectinomycin resistant mutants, a plasmid bearing an IPTG inducible Flp recombinase was mated into transformants and Flp expression was induced to generate in-frame deletions. The plasmid was cured by growing mutants under inducing conditions with 300μg/mL streptomycin. For plasmid expression constructs, a derivative of the pMMB67EH vector with a theophylline controlled riboswitch was used as previously described [55]. For strains made via *sacB* counter selection, a marker-less deletion construct was generated using SOE PCR, and cloned into a pCVD442 suicide vector bearing the *sacB* counter selectable marker and an ampicillin resistance marker via Gibson assembly. LB Agar 10% sucrose plates were used to select for *sacB* loss and recombination of the mutant allele. All constructs were confirmed with DNA sequencing over the region of interest and primer sequences are available upon request.

Real-time quantitative PCR

qPCR experiments were performed as previously described [29]. Briefly, liquid cultures were infected with ICP1 at a Multiplicity of Infection (MOI) of 2.5 at OD_{600}= 0.3. Samples were taken at 0 and 20 minutes post-infection, and boiled before serving as templates for IQ SYBR qPCR reactions. For assays involving induction of *repA*, 2mL cultures were grown with 1.25 μg/mL chloramphenicol for plasmid maintenance and induced for 20 minutes prior to infection using a final concentration of 1.5mM theophylline and 1mM IPTG starting at OD_{600}= 0.17. Primers used for qPCR are listed in Supplementary Table S1.2.

Nanoluciferase reporter assay

Liquid cultures were grown to an OD_{600}= 0.3. Immediately prior to infection and at 4 minute intervals following infection, 100μL of culture infected at an MOI of 2.5 was added to an equal volume of cold methanol. Nanoluciferase production was measured using theNano-Glo® Luciferase Assay System (Promega). The NanoGlo substrate was diluted 50-fold in the NanoGlo buffer and for each sample, 50μL of sample and 50μL of reaction mix were added per well in a black 96 half-well plate. Luminescence was read over 7 minutes at room temperature with 10 seconds shaking between reads. For each biological replicate, the average of 10 reads was used.

Protein purification

E. coli BL21 cells containing a pE-SUMO fusion to *repA* were grown to OD_{600}= 0.5 at 37°C and induced with IPTG to a final concentration of 0.5mM. The cultures were then shifted to 16°C and grown for 24 hours. Cells were centrifuged and resuspended in lysis buffer (50mM Tris-HCl pH 8, 200 mM NaCl, 1mM BME, 0.5% Triton-X, 50mM imidazole, 1 Pierce™ Protease Inhibitor Mini Tablet (Thermo Scientific) and sonicated. Cell debris was removed by centrifugation (29,097 x g for 40 minutes), and the lysate was applied to a nickel resin affinity column (HisPur Ni-NTA Resin, Thermo Scientific). The column was washed with two column volumes of wash buffer (50mM Tris-HCl pH 8, 200mM NaCl, 1mM BME, 50mM imidazole), one column volume of an additional high salt wash (50mM Tris-HCl pH 8, 2M NaCl, 1mM BME, 50mM imidazole) to remove any residual DNA, and then eluted with elution buffer (50mM Tris-HCl pH 8, 200 mM NaCl, 1mM BME, 300mM imidazole). The eluate was dialyzed with sizing buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 1mM Dithiothreitol) in 10K MWCO SnakeSkin dialysis tubing (Thermo Scientific), and the His6-Sumo-tag was cleaved with 1 µL SUMO protease per 100 µg of protein. The eluate was fractionated on a HiLoad 16/60 Superdex 75 size-exclusion column (GE Healthcare) and fractions were analyzed using SDS page. Protein was concentrated using an Amicon Ultra 15mL 3K NMWL centrifugal filter (Millipore Sigma).

Fluorescent Labeling of Oligos

Single stranded oligos were labeled with 5'-TAMRA using the 5' EndTag™ Nucleic Acid Labeling System (Vector Laboratories) following the manufacturer's protocol to label 0.6nM of single stranded probe was labeled with a 5mg/mL solution of tetramethylrhodamine-5-maleimide in DMSO. Following end labeling, the labeled single stranded probe was annealed to its complementary sequence by mixing equimolar concentrations of complementary oligos in water, heating to 65°C for 4 minutes, and then allowing the reaction to return to room temperature. Probe sequence is available in Supplementary Table S1.2.

Electrophoretic mobility shift assays (EMSAs)

80 nM probe was incubated with purified RepA at 30°C for 20 minutes in 20µL reactions with 10mM HEPES pH 7.8, 10% Glycerol, 1µM TCEP, 10mM $MgCl_2$, and 0.4µg poly d(I-C) (Sigma-Aldrich) serving as a nonspecific competitor. The full reaction volume was then loaded onto 8% acrylamide 0.5X Tris-borate gels and ran for 20 minutes at 120V before visualization.

Preparation of phage infection samples for DNA sequencing

A 6mL bacterial culture was infected at OD_{600}= 0.3 with ICP1 at an MOI of 1. At the indicated time points, 1mL was removed from the culture tube and mixed with 1mL ice cold methanol to stop DNA synthesis. These samples were pelleted at 21,694 x g for 2

minutes at 4°C. The methanol was removed through aspiration and the pellet was washed with 1 mL cold
phosphate-buffered saline. Pellets were frozen in liquid nitrogen and stored at -80°C until total DNA was isolated using the QIAGEN DNeasy Blood and Tissue Kit. Sequencing libraries were prepared using NEBNext® Ultra™ II FS DNA Library Prep Kit. Paired-end sequencing (2 × 150 bp) was performed on an Illumina HiSeq4000 (University of California, Berkeley QB3 Core Facility).

DNA-seq reads mapping

Illumina sequencing reads for each timepoint were mapped to the appropriate reference sequence using Bowtie 2 v2.3.4.1 [58] with default settings except for the following: '--end-to-end' and '--very-sensitive'. Mapping files were sorted and indexed with samtools v1.5 [59] and binned with breseq BAM2COV v0.33.0 [60]: ICP1 1000bp, PLE 150bp. Read coverage was normalized by the total number of reads that mapped to the reference. Triplicate experiments were then averaged and plotted with the matplotlib module v3.0.3 in Python [61]. GC skew was calculated over a 1000bp sliding window. For plotting the abundance of a specific genome in a sample, the genome per million (GPM) was calculated in the same manner as the previously described transcripts per million [62].

Protein structure visualization

Protein structure figures were generated using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). Protein structure alignments were generated using the cealign command [63]. The electrostatic distribution was determined and visualized using the PDB2PQR server, and the APBS plugin for PyMOL [64].

Transduction assays

Transduction assays were performed as previously described [29]. Briefly, donors were grown to OD_{600}= 0.3 and infected with ICP1 at an MOI of 5. Cultures were incubated for 5 minutes before being washed with fresh LB to remove unbound phage. The infected cultures were incubated for 30 minutes, and 100μL lysate was added to 100μL of an overnight culture of the recipient strain. The mixture was incubated for 1 hour at 37°C with aeration before plating on selective media.

Efficiency of Center of Infection assay

Cultures were grown to an OD_{600}= 0.3, at which time they were infected at an MOI of 0.1 and incubated for 7.5 minutes to allow phage attachment before being diluted 2500-fold in warm LB. 500μL of this dilution was collected and treated with 20μL chloroform to enumerate phage input. The diluted cultures were diluted further, 10 and 100-fold into two additional tubes containing 2mL LB. The dilution series of infected cells was then returned to the incubator. Infected cells were collected at 35 minutes post initial

infection. A plaque assay was performed by adding the infected cells to PLE(-) cells to measure the center of infection for the strains of interest.

RESULTS

PLE alters and diminishes ICP1 replication

PLE was previously shown to replicate to high copy during ICP1 infection and reduce ICP1 DNA replication compared to a PLE(-) control [29] (Fig 1.1), however, these results were obtained through qPCR and only assessed a single ~100bp target sequence. To obtain a more complete understanding of PLE replication dynamics and the PLE's impact on ICP1 replication kinetics, we performed deep sequencing of total DNA during an ICP1 infection time course using PLE(-) and PLE(+) *V. cholerae*. ICP1 produces new progeny virions by 20 minutes post-infection in PLE(-) cultures, and PLE(+) cultures lyse 20 minutes post-infection [29], therefore to evaluate total DNA content in infected cells at early, middle and late time points (while avoiding potential DNA loss due to lysis), we collected samples at 4, 8, 12 and 16 minutes post-infection. Total DNA from samples at each time point was sequenced on an Illumina HiSeq and the resulting sequencing reads were mapped against the *V. cholerae,* ICP1, and PLE genomes. Consistent with the anticipated rapid kinetics of ICP1 infection in PLE(-) *V. cholerae*, the abundance of ICP1 reads increased within 8 minutes post-infection and ICP1 DNA comprised roughly half of the total DNA content by 16 minutes post-infection (Supplementary Table S1.3). To account for the relatively small size of the ICP1 genome compared to the *V. cholerae* chromosomes, we normalized the reads mapped per element to element length and the total reads per sample to determine the genomes per million (GPM) of each entity in the samples. Prior to infection, the GPM for the *V. cholerae* large chromosome is higher than that for the small chromosome (Fig 1.2A), consistent with previous studies showing that replication of the small chromosome initiates after the large chromosome, leading to roughly synchronous termination [65], and that replication initiation of the small chromosome requires duplication of certain loci in the large chromosome [66]. Following infection, ICP1 DNA replication robustly overtakes the cell and phage genomes are more abundant than copies of the *V. cholerae* large and small chromosomes by 12 minutes post-infection (Fig 1.2A). In contrast, ICP1 DNA replication is less robust in the presence of PLE. Specifically, the proportional abundance of ICP1 DNA is relatively unchanged at 4 and 8 minutes post-infection of PLE(+) cells, but ICP1 DNA replication begins to dramatically lag by 12 minutes post-infection compared to PLE(-) infection (Supplementary Table S1.3). In the PLE(-) condition, ICP1 relative reads abundance doubles from roughly one quarter to one half of total reads between 12 and 16 minutes post-infection, while in the PLE(+) condition ICP1 abundance increases very little between 12 and 16 minutes post-infection (Supplementary Table S1.3). The defect observed in ICP1 replication correlates with PLE's own robust replication. By 8 minutes post-infection, PLE is already the most abundant element in terms of copy number (Fig 1.2B). Between 8 and 16 minutes post-infection, the abundance of PLE DNA grows to comprise approximately 19% of total reads, overtaking ICP1 in total DNA (Supplementary Table S1.3). In terms

of genome copy at 16 minutes post-infection, PLE outnumbers ICP1 approximately eight-fold (Fig 1.2B). The temporal dynamics of PLE and ICP1 DNA replication support the notion that interference of ICP1 replication may be linked to the PLE's own replication.

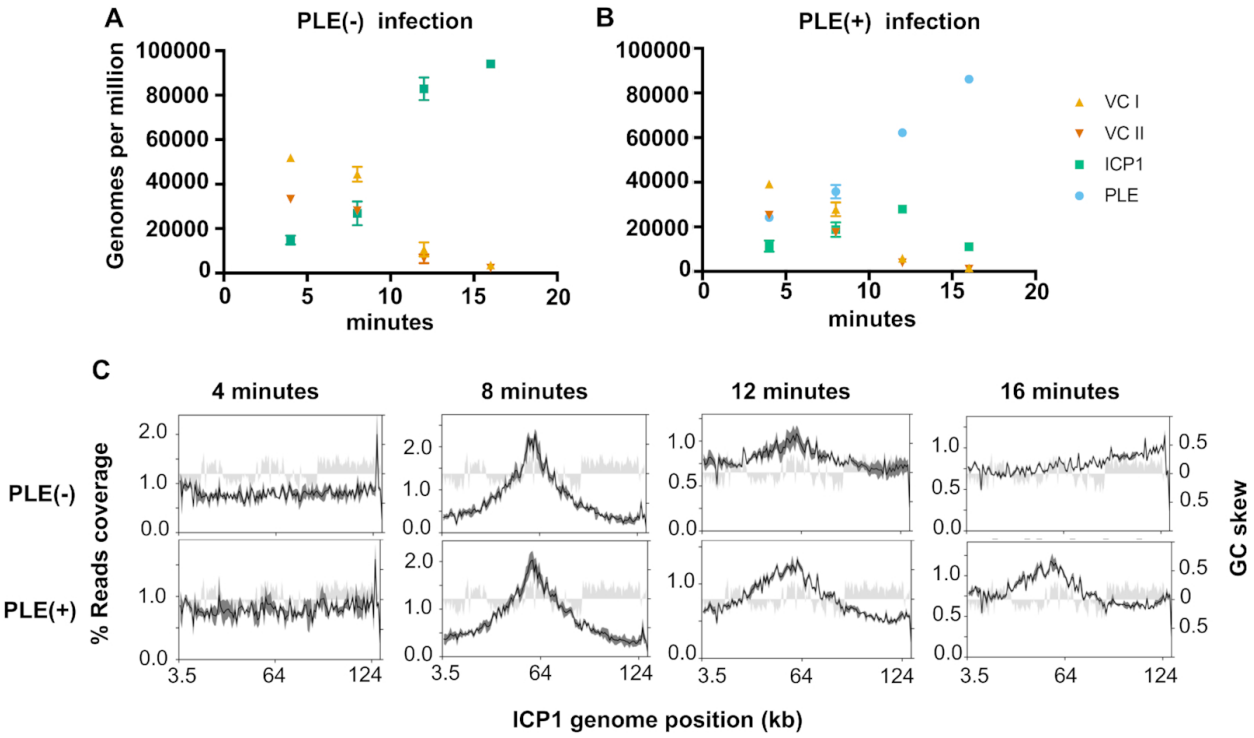


Figure 1.2. PLE robustly replicates following infection while altering ICP1 replication. (A and B) Genomes per million (GPM) of total DNA mapping to the V. cholerae large (VC I) and small (VC II) chromosomes, ICP1, and the PLE across an infection time course in PLE(−) (A) and PLE(+) (B) V. cholerae. Samples were taken at 4, 8, 12 and 16 min post-infection, data show the average and standard deviation of three independent experiments. (C) Percent reads coverage plots across the ICP1 genome during PLE(−) (top) and PLE(+) (bottom) infection. For each time point, the percent reads coverage across the genome for three biological replicates was determined. The average percent reads coverage is shown as a black line, while standard deviation appears as dark gray shading around the line. The GC skew (right axis) is shown as light gray shading.

In addition to monitoring the relative changes in abundance of discrete genetic elements during phage infection, we evaluated the profiles of sequence coverage across ICP1 and PLE genomes (Fig 1.2C, Supplementary Fig 1.1). While the distribution of reads across the ICP1 genome was similar in PLE(+) and PLE(-) conditions at 8 minutes post-infection, ICP1's coverage profile was markedly different at 12 and 16 minutes post-infection between the two conditions (Fig 1.2C). At 8 minutes post-infection, a peak in ICP1 reads can be seen near the 60kb position, and reads abundance decreases with increasing distance from that point. The observed pattern in ICP1, which is present in both PLE(+) and PLE(-) infections, is consistent with the predicted coverage of an element that replicates bidirectionally through theta-replication from a single origin of replication (ori) [67]. At 16 minutes post-infection in the PLE(-) condition the peak reads abundance shifts to one end of ICP1's genome (Fig 1.2C). These results suggest activation of an additional ICP1 ori late in infection. Additionally, the distribution of reads decreases gradually in the upstream direction from the peak, and sharply drops

downstream of the peak. Such a distribution is suggestive of a rolling circle mode of replication [67], which is consistent with a number of phages that are known to transition to rolling circle late in infection [54]. By contrast, at 16 minutes post-infection in PLE(+) *V. cholerae*, the profile of ICP1 reads more strongly resembled the profile at 8 minutes post-infection than it did to the coverage profile 16 minutes post-infection in the PLE(-) condition. The change in ICP1 reads distribution suggests that PLE might alter ICP1 replication origin choice and impair the progression from theta to rolling circle replication.

Intriguingly, the reads peak at near the end of ICP1's annotated genome was prominently visible at 4 minutes post infection, before ICP1 replication has taken place (Fig 1.2C). We speculated that this reads peak corresponds with the terminus of infecting ICP1 particles prior to genome circularization as it has previously been established that termini can lead to sequencing biases following DNA library preparation [68]. We found that this reads bias was also present in DNA from purified phage particles (Supplementary Fig 1.2). PhageTerm, which identifies phage termini and packaging methods [68], identified this reads peak as a packaging (pac) site and predicted a headful packaging mechanism for ICP1. The terminus is located in a 1.3kb orfless space between *gp1* and *gp2* (Supplementary Fig1.2A and B). PhageTerm predicts the location of the pac site at 431bp on the annotated (+) strand, and 891bp on the annotated (-) strand (Supplementary Fig 1.2C). The loss of this peak by 8 minutes post-infection (Fig 1.2C) likely reflects circularization of the phage genome following cell entry. Together, ICP1's changing coverage profile over the course of infection, and PhageTerm analysis of phage particle DNA suggests that rolling circle initiation and genome packaging may be linked for ICP1. That the shift in ICP1's coverage profile was profoundly reduced in the PLE(+) background suggests that PLE interferes with the rolling circle mode of ICP1 replication, potentially preventing the switch from theta replication. This interference of rolling circle replication may perturb later steps (i.e. DNA packaging) necessary for ICP1 to complete its life cycle.

PLE encodes its own replication initiator, but does not replicate autonomously

To better understand the relationship between PLE and ICP1 DNA replication, we next sought to identify the constituents of the PLE replicon. The PLE genome is 18kb and organized into multiple predicted gene clusters (Fig 1.3A). Between PLE *orf5* and *orf7*, is a 2.7kb non-coding region (NCR) which has four repeat sequences (Fig 1.3B, Supplementary Fig 1.3). Frequently, repetitive sequences serve as binding sites for replication machinery at phage and plasmid origins of replication [69]. Within bacterial genomes, there is also a bias for coding sequence in the leading strand [70], and this is consistent with an ori being between divergently transcribed operons. These features led us to hypothesize that the PLE NCR serves a function in replication. This was further evidenced by PLE's replication reads profile, which showed a peak approximately 1kb upstream of *orf7* at 8 minutes post-infection, suggesting that the PLE ori is located in the NCR (Fig 1.3A). To test if the NCR contained sequence necessary for PLE replication, we generated three strains designated NCR1, NCR2 and NCR3 that each possessed a 0.5-1kb deletion within the NCR excluding predicted promoters for

orf5 and *orf7* (Fig 1.3B). Following ICP1 infection we found that NCR1 and NCR2 were dispensable for PLE replication, however, NCR3, containing repeat sequences 3 and 4, was necessary for replication (Fig 1.3C). This suggested that the PLE ori was contained within NCR3, and that repeat 3 and/or repeat 4 may be involved in replisome recruitment. We next sought to determine whether any predicted PLE open reading frames (ORFs) are necessary for PLE replication. We began by screening PLE gene cluster knockouts [55] and we observed that one gene cluster, containing *orf7* through *orf14,* was necessary for PLE replication (Fig 1.3D). To identify the ORF responsible, we constructed individual gene knockouts within this cluster and screened for replication defects during ICP1 infection. We found that a single open reading frame, *orf11* (https://www.ncbi.nlm.nih.gov/protein/AGG09405.1/), was necessary for PLE replication (Fig 1.3E). Given the requirement of *orf11* for PLE replication, and further analyses supporting its designation as a replication initiation protein (discussed below) we designated *orf11* as *repA.*

We next wanted to check if ICP1 induces expression of *repA*. We tested this by infecting a Δ*repA* reporter strain with nanoluciferase under the native *repA* promoter. Samples from infected cultures and uninfected controls were taken just prior to infection and at 4 minute intervals following infection. The luminescence activity of the infected strain was noticeably higher at 8 minutes post infection, and continued to climb at 12 and 16 minutes, confirming that ICP1 infection activates expression of *repA* (Supplementary Fig 1.4).

In both P4 and SaPIs, satellite replication is autonomous following the satellite's transcriptional activation by the helper phage [49,71]. Having determined that PLE-encoded *repA* is induced upon ICP1 infection and necessary for PLE replication, we sought to elucidate if expression of *repA* was sufficient to drive autonomous replication of PLE. We complemented PLE *ΔrepA* with ectopically expressed *repA* and measured PLE copy number increase in the presence and absence of phage. RepA expression was able to drive PLE replication, but only in cells infected by ICP1 (Fig 1.4A). Our infected uninduced culture exhibited a low level of PLE replication, presumably due to leakiness of the expression construct. Consistent with this, ICP1 was unable to drive replication of the Δ*repA* PLE complemented with an induced empty vector control (Supplementary Fig 1.5). This result shows that RepA is necessary for PLE replication.

To rule out the possibility that PLE replication requires additional PLE genes activated by ICP1 that may have been missed in our genetic screen due to redundancy, we next set out to define the minimal unit required for PLE replication. Previous work showed that ICP1 infection triggers excision of a 'miniPLE', consisting of the PLE encoded integrase together with a kanamycin resistance marker flanked by the PLE attachment (att)-sites [55]. We built on this existing platform and constructed a 'midiPLE' which differs only by the presence of the NCR containing the PLE ori on the self-excising miniPLE (Fig 1.4B). The midiPLE replicated to a substantial level that was dependent on ICP1 infection and *repA* expression, replicating to about 40% the copy number of a wild-type control (Fig 1.4C). This result confirmed that *repA* and the PLE non-coding region are sufficient to drive PLE replication, but only during phage infection.

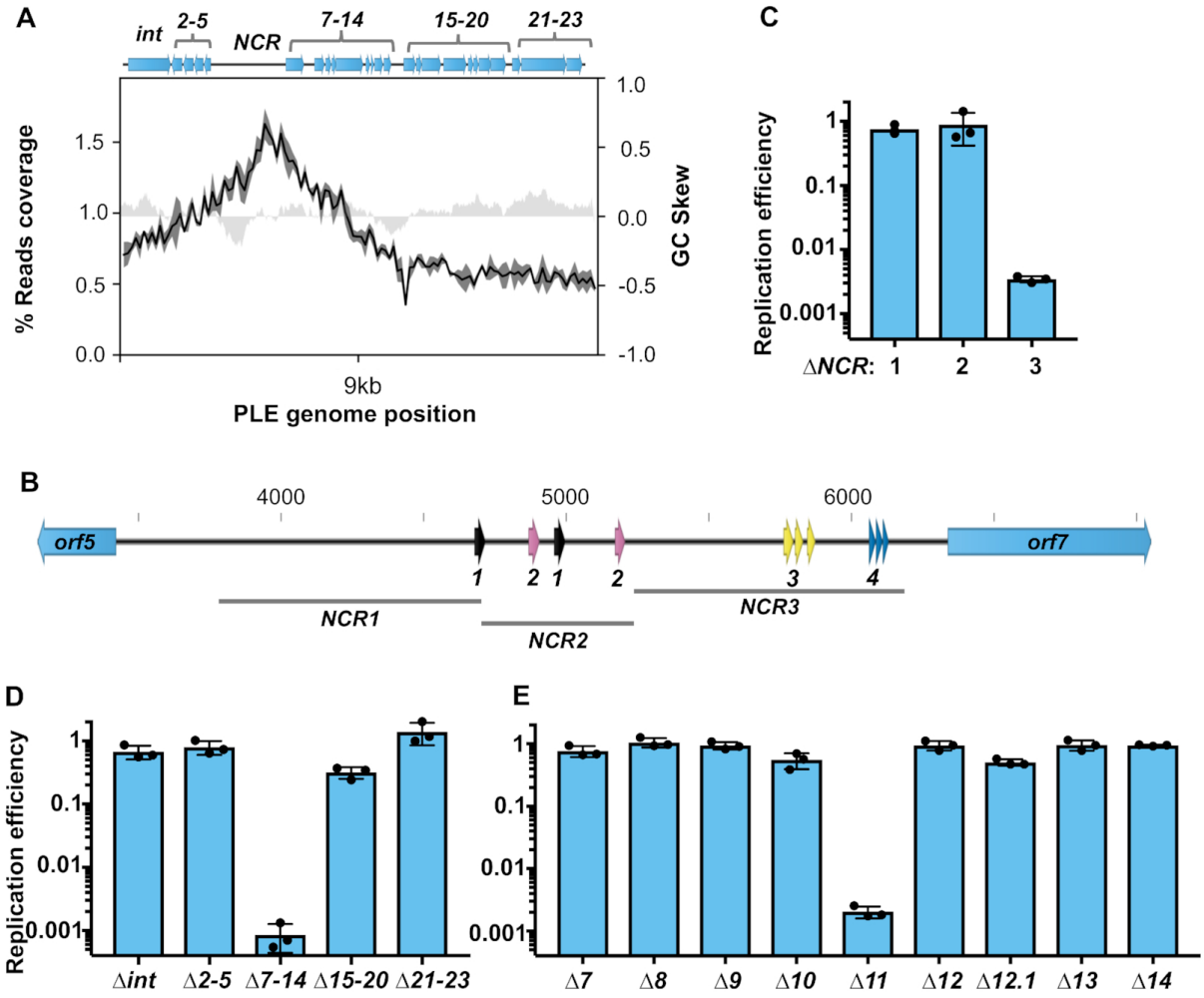


Figure 1.3. A single PLE-encoded ORF and a noncoding region are necessary for PLE replication. (A) A representation of the PLE genome (top) with average reads coverage of PLE 8 min post ICP1 infection plotted below. The percent reads coverage was determined for three biological replicates and is shown as a black line, while standard deviation appears as dark gray shading around the line. The GC skew (right axis) is plotted as light gray shading. Gene clusters mutated for analysis are labelled. (B) PLE's noncoding region (NCR) between orf5 and orf7, with repeat sequences shown as arrows. Repeat sequences share colors for each repeat type, and are designated as repeats 1, 2, 3 or 4. Regions of the NCR deleted for analysis in (C) are shown. Panels C–E: replication of PLE mutants 20 min post-infection with ICP1 as assessed by qPCR. Replication efficiency is relative to a wild-type PLE control. (C) Replication of ΔNCR mutants. (D) Replication of PLE gene cluster knockouts. (E) Replication of individual gene knockouts of the ORFs contained in cluster 7–14.

To determine if the capacity to drive PLE replication is unique to ICP1, we also tested PLE replication during infection by ICP3, an unrelated T7-like phage [26]. Being a T7-like phage, ICP3 encodes a DNA polymerase and helicase-primase belonging to the same families as those of ICP1. To avoid the midiPLE being potentially degraded along with the host chromosome by ICP3, we complemented Δ*repA* PLE with the ICP1-encoded recombination directionality factor PexA to stimulate PLE excision [55], as well as RepA. Due to shared promoters and the toxicity of inducing PexA, we were unable to induce expression of RepA prior to infection, precluding high levels of replication. Nevertheless, during infection by ICP1 under these conditions, the PLE replicated

upwards of 50-fold (Supplementary Fig 1.6). By contrast, the ICP3 infected cultures and uninfected controls did not show any evidence of PLE replication, suggesting that PLE replication may require components uniquely encoded by ICP1.

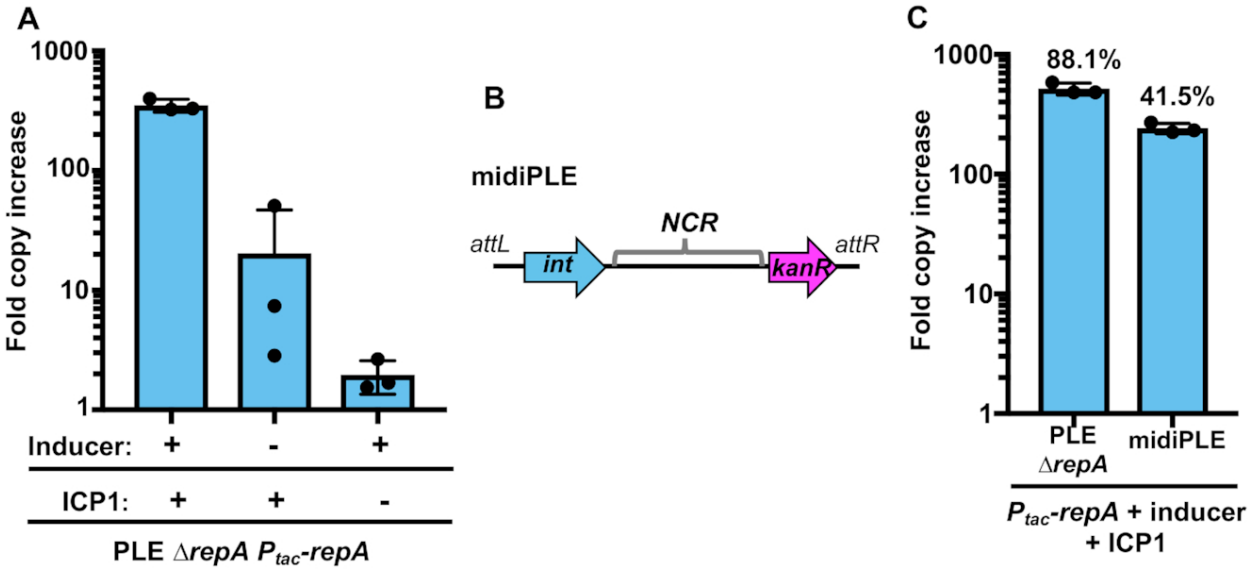


Figure 1.4. RepA drives PLE replication in the presence of ICP1. (A) RepA complementation of PLE ΔrepA as assessed by qPCR. PLE fold copy increase 20 min post-infection is shown in different combinations with ICP1 and the inducer of the complementation construct. (B) A diagram of the midiPLE construct used to assess the minimal requirements for PLE replication (not to scale). Attachment sites, the PLE integrase, and the noncoding region (NCR) are present along with a kanamycin resistance gene (kanR). (C) Replication of a RepA complemented ΔrepA strain and midiPLE 20 min post ICP1 infection. The replication of these strains was compared to a wild-type PLE control, and the relative replication is displayed as a percentage above the bars.

The PLE encoded replication protein RepA resembles Gram-positive plasmid initiation factors

Although the structure and function of PLE's RepA has not been previously elucidated, the x-ray crystal structure of the N-terminal domain (NTD) of RepA (RepA-NTD) has been solved and deposited in the Protein Data Bank (PDB ID: 4RO3) (Fig 1.5A). While primary sequence similarity (using BLASTP) is not evident, using Dali [72], we found that PLE RepA-NTD has substantial structural similarity to the pKS41 and pTZ6162 plasmid RepA proteins from *Staphylococcus aureus*, as well as more distant similarity to the replication protein DnaD from *Bacillus subtilis* (PDB ID: 4PTA, 4PT7, and 2v79). Both of the *S. aureus* RepA proteins serve as replication initiators for plasmids coding for multidrug resistance and belong to the RepA_N family of plasmid replication proteins. The RepA_N protein family is comprised mostly of initiation factors for theta-replication of plasmids found mainly in the Firmicutes [73,74]. This protein family is named for the conservation of the NTD which structurally resembles the NTD of the Gram-positive primosome component DnaD [75]. In the RepA_N family, the NTD mediates DNA binding, while the C-terminal domain (CTD) of these proteins appear to be specific to host genus, and may perform host specific functions [73]. An HHPRED [76] did not detect any substantial similarities to RepA's CTD (expect >1 for al hits).

The PLE RepA-NTD structure aligns well with the crystal structures of the NTDs from *S. aureus* pTZ6162 and pSK41 RepA initiation factors, highlighting shared tertiary structure (Fig 1.5B, Supplementary Fig 1.7). Notably, all three proteins crystallized as dimers with monomers in the same orientation, suggesting a conserved dimer interface, and the potential for a conserved method for DNA binding. A crystal structure for the pTZ6162 RepA-NTD dimer bound to its cognate iteron dsDNA sequence has also been solved (PDB ID: 5kbj) [75],and we were able to align the PLE RepA-NTD to this structure (Fig 1.5C). In the original pTZ6162 structure, the surface of the protein that is bound to the DNA is electropositive [75], consistent with binding activity for the electronegative dsDNA sugar-phosphate backbone. The corresponding surface in the PLE RepA-NTD structure is electropositive as well, suggesting conserved DNA binding activity (Fig 1.5D). An electropositive DNA binding interface is also observed in the pSK41 RepA structure suggesting maintenance of a shared DNA binding region among these proteins (Supplementary Fig 1.8A-C). Notably, the DnaD-NTD is less electropositive along the corresponding surface, which is expected since the DnaD-NTD, despite its structural similarity to RepA_N-NTDs, does not bind DNA (Supplementary Fig 1.8D) [73]. Given PLE RepA's structural similarity to the RepA_N family, its similar electrostatic profile, and its shared role as a replication factor for a mobile genetic element, we conclude that PLE RepA belongs to the RepA_N protein family.

Like other replication initiation factors [69], RepA_N family proteins bind to repetitive iteron sequences at their cognate ori. Most characterized RepA_iterons are semi-palindromic direct repeats, containing inverted repeats that converge on a poly-A tract [73] (Fig 1.5E). These same sequence features are apparent in repeat 3 and repeat 4 in NCR3 (Fig 1.5E), which is necessary for PLE replication (Fig 1.3B). The iterons for pTZ2162 and pSK41 have repeats of 9 and 8bp respectively. Repeat 3 in PLE has inverted repeats that are longer at 13bp, while those in repeat 4 are only 3bp long. Most characterized RepA_N iteron inverted repeats are at least 5bp long, but obvious inverted repeats are not always discernible [73]. To determine if PLE's RepA is capable of binding to the repetitive sequences in NCR3, we purified RepA and assessed binding to repeat 3 and repeat 4 through an EMSA. When RepA was titrated into our reactions we observed that the repeat 3 probe, but not the repeat 4 probe was shifted on the gel, confirming that RepA binds the repeat 3 sequence (Fig 1.6B). Additional genetic analysis showed that the repeat 3 sequence was necessary for PLE replication, but the repeat 4 sequence was not, further supporting our conclusion that repeat 3 serves as the iteron sequence in the PLE ori (Fig 1.6C).

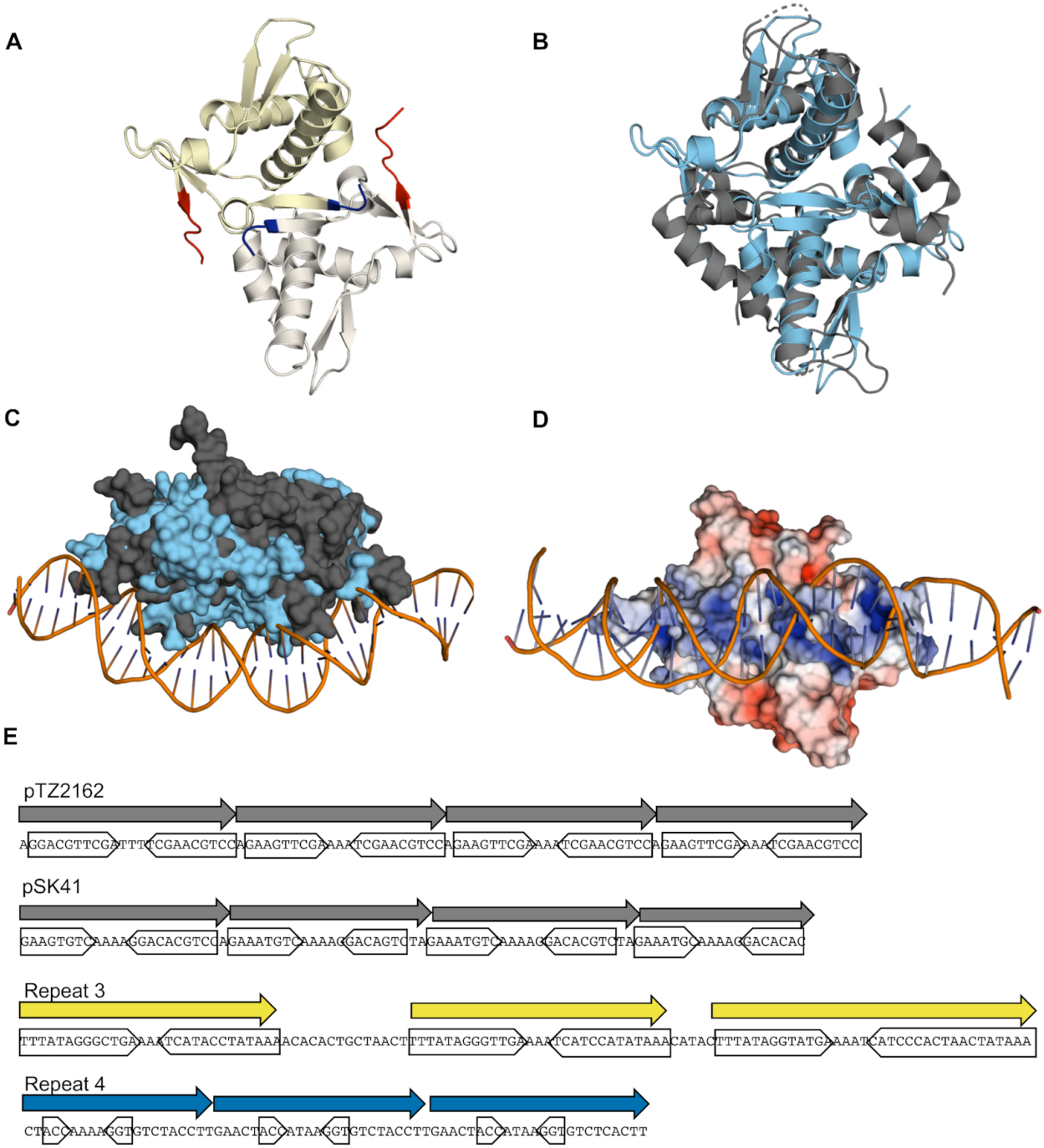


Figure 1.5. PLE RepA is a RepA_N family protein. (A) Cartoon representation of the PLE RepA-NTD dimer. Monomers are differently colored in yellow and white. The N and C termini of the monomers are colored blue and red respectively. (B) Alignment of the NTD dimers of PLE RepA and pTZ2162 RepA in light blue and dark grey, respectively, depicted in cartoon representations (RMSD = 4.527 over 176 residues). (C) Surface view of PLE RepA-NTD dimer in light blue aligned with pTZ2162 RepA-NTD dimer in dark gray bound to substrate DNA. (D) Electrostatic potential map, turned 90 degrees as (C), of PLE RepA-NTD dimer aligned to the pTZ2162 RepA-dsDNA bound structure. Positive (blue) and negative (red) charges are indicated on the surface. (E) Binding iterons for the RepA initiators of pTZ2162 and pSK41 are shown alongside repetitive sequences found in the putative PLE origin of replication. Direct repeats are denoted with an arrow, while the sequence comprising inverted sub-repeats is boxed. Sequence for the minus strand for PLE is shown to make the central poly-A tract apparent.

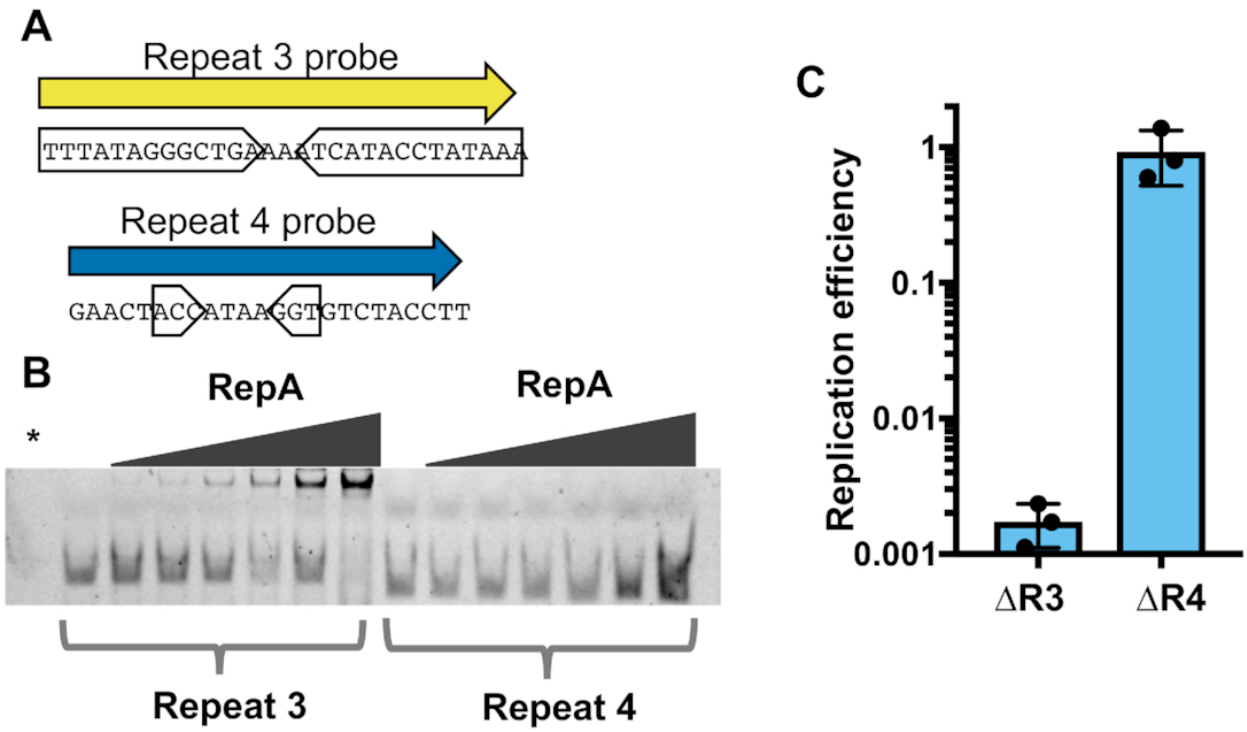


Figure 1.6. Repeat 3 serves as the PLE ori iteron sequence. (A) The nucleotide sequences of 5′ fluorescently labelled dsDNA probes used to test RepA specificity. Probe sequence was derived from the repetitive sequences found in PLE noncoding region 3 (NCR3). Inverted sub-repeats are boxed. (B) An electrophoretic mobility shift assay using the probes in (A). The * denotes a RepA(+) no DNA control. Additional replicates of this experiment are shown in Supplementary Figure S1.9. (C) The replication of PLE mutants with deletions spanning the repeat 3 (R3) and repeat 4 (R4) regions 20 min post-infection with ICP1 as assessed by qPCR. Replication efficiency is relative to a wild-type PLE control.

Non-replicating PLE alters ICP1 replication dynamics without lowering ICP1 genome copy

Having identified the necessary components of the PLE replicon, we sought to assess the importance of PLE replication for the PLE's life cycle and anti-phage activity. Following excision and replication, PLE can be transduced to recipient *V. cholerae* cells [29]. We hypothesized that PLE replication would be necessary for its transduction, therefore we performed transduction assays, comparing Δ*repA* PLE complemented with either *repA* or an empty vector control. As expected, PLE transduction was below the limit of detection in Δ*repA* PLE complemented with an empty vector control (Supplementary Fig 1.10), and the transduction defect for Δ*repA* PLE could be complemented by *repA in trans*, restoring PLE transduction to levels near those of wild-type (WT) PLE (3.8×10^4 units/mL, [29]).

The finding that high PLE copy is needed to facilitate PLE transduction is intuitive, but under these laboratory conditions PLE produces fewer than one transducing unit per infected cell, despite PLE's robust replication [29]. This lead us to question if PLE replication contributes to PLE's anti-phage activity. ICP1 replication is reduced in a PLE(+) infection (Fig 1.2). A potential mechanism of PLE impairment of ICP1 replication could be through the consumption of replication resources. Robust PLE replication

might exhaust dNTP pools, and since PLE only replicates during ICP1 infection, PLE might also competitively restrict ICP1's access to its own replisome. Therefore, we next tested ICP1 replication in non-replicating PLE strains using qPCR, and observed that ICP1 replication was restored to the levels seen in PLE(-) infection conditions (Fig 1.7A). This restoration led us to question if midiPLE replication could impair ICP1 replication simply by using up replication resources. However, during ectopic expression of *repA*, midiPLE did not impair ICP1 replication, while *ΔrepA* PLE did (Fig 1.7B). Consistent with this result, we did not observe any defect in ICP1 plaque formation on *V. cholerae* harboring a replicating midiPLE (Supplementary Fig 1.11). This suggests that PLE replication reduces ICP1 copy through an independent mechanism, which may be dependent on PLE gene dosage increase, or by reaching a level of replication not achievable with the midiPLE.

The roughly four-fold decrease in ICP1 replication that occurs in PLE(+) cultures would not likely be sufficient for the complete restriction of ICP1 that is observed [29], but is likely to be a contributing mechanism. To investigate this, we performed ICP1 plaque assays on non-replicating PLE mutant hosts. The PLE *ΔrepA* and *Δori* mutants still blocked plaque formation (data not shown), however, the mutants were more susceptible to ICP1 than wild-type PLE as some small plaques were visible when high phage concentrations were added. The small size of these plaques made quantification difficult and less reproducible than desired. Therefore, we quantified ICP1's efficiency of center of infection (EOCI) on these non-replicating PLE mutants (Fig 1.7C). Consistent with previous observations [29] virtually no phage were produced from wild-type PLE(+) *V. cholerae*. We did, however, observe an intermediate EOCI on the non-replicating PLE strains. Unexpectedly, a double knockout of both the iteron sequence and *repA* permitted less phage production than each individual knockout. This is peculiar, but could make sense if PLE replication has downstream regulatory effects on PLE activity. Since the iterons and RepA are interacting partners, removing both may allow PLE to bypass any regulatory activities either may have rather than becoming arrested at failed replication initiation. These results illustrate the difficulty of teasing apart direct and downstream effects of PLE replication. Nevertheless, each of our non-replicating PLE mutants had some restoration of phage production, demonstrating that although PLE replication is not necessary for PLE mediated anti-phage activity, PLE replication bolsters or acts synergistically with other PLE encoded anti-phage activities.

We observed that PLE replication decreases the level of ICP1 replication and coverage profiles suggested that PLE inhibits ICP1's transition to rolling circle replication (Fig 1.2C). Since qPCR experiments indicated that the level of ICP1 replication was restored when PLE replication was abolished, we last wanted to determine if ICP1's change in replication mode was also restored when PLE replication was abolished. Therefore we performed deep sequencing of total DNA during an ICP1 infection time course in PLE *ΔrepA* and quantified and mapped coverage from 8 and 16 minute post-infection samples. As expected and consistent with the qPCR results, the relative abundance and GPM of ICP1 did not differ from what we saw for the PLE(-) infection conditions (Fig 1.8A, Supplementary Table 1.4). As seen before, the coverage profile of ICP1 at 8 minutes post-infection shows that ICP1 uses a bidirectional mode of replication at that

time point. Interestingly, while loss of PLE replication restored ICP1 copy, abundance across the ICP1 genome matched neither the PLE(-) nor wild-type PLE(+) culture conditions at 16 minutes post-infection (Fig 1.8B). The highest abundance of reads was shifted near to the end of ICP1's annotated genome, as in the PLE(-) infection, but the gradual decrease in reads from this point was bidirectional rather than unidirectional (Fig 1.8C). This reveals that PLE has some capacity to act on ICP1 replication even when the PLE is not replicating, and suggests that PLE may prevent linearization of ICP1's genome.

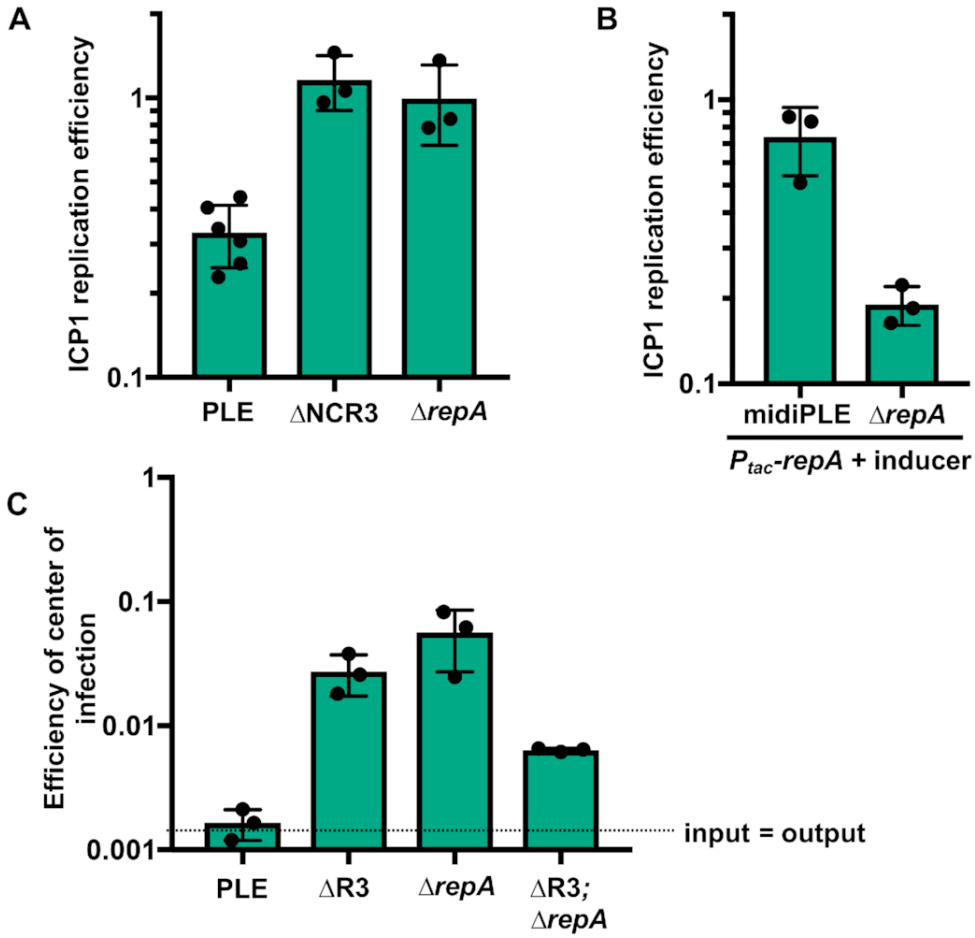


Figure 1.7. Loss of replication impairs PLE anti-phage activity. (A) ICP1 replication in wild-type and mutant PLE(+) strains as assessed by qPCR. Replication efficiency is relative to ICP1 infection of PLE(−) V. cholerae 20 min post-infection. (B) Replication of ICP1 as assessed by qPCR in RepA complemented midiPLE and ΔrepA PLE infection relative to an un-complemented midiPLE control. (C) Efficiency of center of infection (EOCI) for ICP1 on wild-type PLE and non-replicating PLE mutant hosts. ECOI is relative to a PLE(−) permissive control strain. The dashed line indicates the threshold at which the number of output phage is equal to the number of input phage. Above the dashed line output has a larger value, below the dashed line input has a larger value.

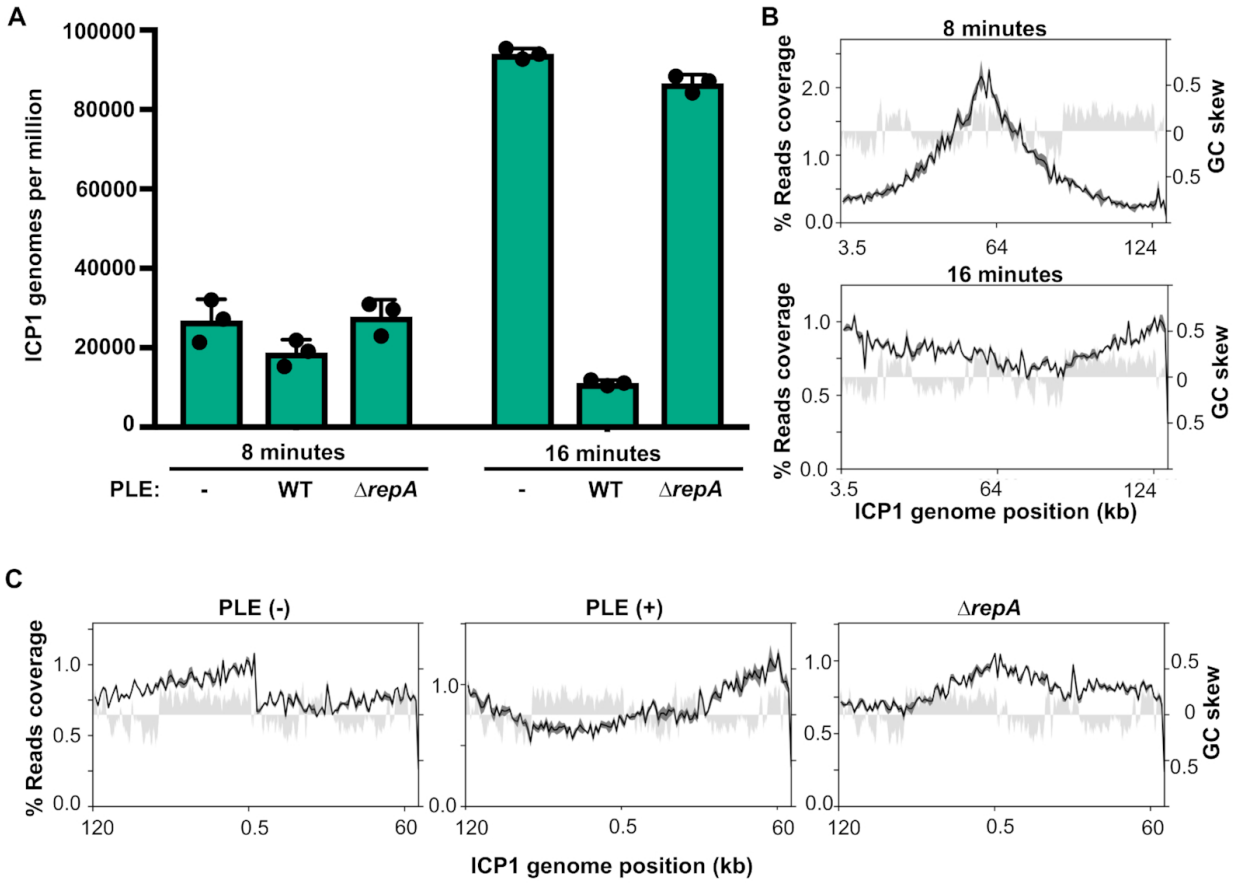


Figure 1.8. Non-replicating PLE still alters ICP1's replication profile. (A) Genomes per million of ICP1 in PLE(−), PLE(+), and PLE ΔrepA cultures at 8 and 16 min post-infection. Values shown are the means of three biological replicates. (B) Percent reads coverage profile of ICP1's genome in ΔrepA PLE infection at 8 min (top) and 16 min (bottom) post-infection. (C) Percent reads coverage profile of ICP1's genome in PLE (−), PLE (+), and ΔrepA PLE hosts at 16 min post infection. The ICP1 genome has been rotated so that it is centered around the putative rolling circle replication origin. For each reads profile plot in (B) and (C), the -average percent reads coverage across the genome for three biological replicates is shown as a black line, while standard deviation appears as dark gray shading around the line. The GC skew (right axis) is plotted as light gray shading.

DISCUSSION

Here, we have identified the key constituents of PLE replication and evaluated their importance for ICP1 restriction. Since PLE's genetic material outnumbers ICP1's by 16 minutes post-infection, it is easy to imagine that reduced ICP1 copy, coupled with the presence of a highly abundant competing genome could severely hamper ICP1 packaging. Still, neither a decrease in ICP1 copies, nor a high abundance of PLE copies is necessary for PLE's anti-phage activity, indicating that PLE has other mechanisms for restricting ICP1. Our results indicate that one of these mechanisms may still be centered around ICP1's replication even if it does not decrease the overall ICP1 copy. The coverage profiles of ICP1 suggests that it undergoes a transition in replication mode from bidirectional replication to rolling circle replication, and that PLE impedes this transition even when PLE is unable to replicate (Supplementary Fig 1.12). A number of well characterized phages are known to transition from bidirectional theta to rolling circle

replication over the course of infection [54]. This transition linearizes and concatenates the phage genome, and concatemeric DNA serves as the packaging substrate for most tailed phage [77]. If PLE *ΔrepA* still prevents ICP1 from replicating via a rolling circle mechanism, this could severely impair ICP1's ability to package its genome. PhageTerm analysis suggests that the ICP1 pac site is proximal to where we predict the rolling circle replication origin, potentially linking rolling circle replication and genome packaging. It is conceivable that the blunt terminus generated by the first round of rolling circle replication could act as a recognition site for the ICP1 terminase, which would then package the concatemeric genome in a headful fashion. Additionally, if a loss of genome linearization is not sufficient to prevent ICP1 particle production, it could act synergistically with other anti-packaging mechanisms such as the capsid hijacking observed in SaPIs and P4.

The precise relationship between ICP1's and the midiPLE's (and by extension PLE's) DNA replication remains unclear. Specifically, it is unclear if replication of the midiPLE does not interfere with ICP1's replication because the midiPLE does not replicate to the same level as PLE, or if the midiPLE is unable to reach as high of a copy level because ICP1 replication is unperturbed. Further work will be needed to identify factors that act on ICP1 replication without impacting PLE copy.

PLE's lack of autonomous replication is a striking contrast to previously characterized bacteriophage satellites. PLE's reliance on ICP1 for replication and not just activation of gene expression indicates that PLE parasitizes ICP1-encoded gene products for replication. Recent work further demonstrates that PLE parasitizes ICP1 proteins for replication as it was found that PLE replication cannot proceed during infection by ICP1 mutants lacking an SF1B-type helicase [78]. ICP1 isolates encode one of two SF1B-type helicases (*helA* or *helB*) in syntenic loci. Despite sharing only 24% amino acid identity, each of these helicases can be exploited for PLE replication during infection. Surprisingly, Δ*helA* and Δ*helB* phage can drive PLE replication when complemented by Dda, an even more distantly related SF1B-type helicase encoded by phage T4. This is especially remarkable given that ICP1's core replisome does not resemble those of T4-like phage [52,54,79]. The sequence diversity of these SF1B-type helicases, and the unrelatedness of their cognate phages' replisomes suggests that it may be their enzymatic activity, rather than a protein binding affinity, that makes the helicases necessary for PLE replication

Questions remain for understanding PLE replication: specifically, what replication machinery PLE is recruiting to the ori and how that recruitment occurs. Replication proteins for SaPIs and P4 possess primase and helicase activity [49–51] but PLE RepA is not predicted to possess either. How RepA_N proteins initiate replication remains unknown, and given that the C-terminal domain of these proteins are genera specific, it is likely that RepA_N proteins in different bacteria will recruit different machinery and may initiate replication through separate mechanisms. PLE's need for an SF1B-type helicase for replication[78] offers some potential clues into how RepA initiates PLE replication. Though the physiological role of Dda in T4 as well as the SF1B-type helicases in ICP1 remains elusive, Dda has been implicated in loading one of T4's

origins of replication, and may be involved in modulation of recombination [54,80]. Recombination is coupled to replication initiation during processes such as the restarting of stalled replication forks and the recombination dependent replication carried out by T4 phages [54,80]. Rather than recruit ICP1 replication machinery through direct interactions, PLE RepA might be restructuring the PLE origin so that it resembles damaged replication intermediates. ICP1's SF1B-type helicase may then process the origin so that the ICP1 replisome can load onto that site. Using DNA repair processes as a backdoor for replisome loading could explain how PLE is able to use diverse SF1B-type helicases to replicate. Future work will be needed to explore this possibility.

It is surprising that PLE encodes a RepA_N family initiator given their rarity among Gram-negative bacteria. Of the 742 RepA_N family proteins annotated in the pfam database, 723 belong either to Firmicutes species or bacteriophage that infect them (Pfam: Family:RepA_N (PF06970)). Only two RepA_N family sequences have been previously identified in proteobacteria, both of them in the group Burkholderiales. Interestingly, RepA is not the only PLE encoded gene that belongs to a family that is underrepresented in Gram-negative bacteria. The PLE integrase responsible for excision and integration into the host chromosome is a large serine recombinase [55] another protein type rarely found in Gram-negative bacteria and phage [81]. Though PLEs lack any detectable homology to other known satellites, the presence of a large serine recombinase and RepA_N initiator in PLEs raises the possibility of recent inter-phyla gene transfer or deep evolutionary roots for PLEs.

Previously, it was noted that the chromosomally encoded RepA_N family proteins are linked to tyrosine and serine recombinases [73]. The authors speculated that these genes were located on conjugative transposons and that the RepA_N had acquired new activities to facilitate transfer, since conjugative transposons, unlike plasmids or phages, do not need to replicate independently of the chromosome. An equally plausible explanation is that these recombinases and RepA_N genes are encoded by cryptic bacteriophage satellites. Supporting this possibility, a *Clostridium difficile* conjugative transposon encoding both a serine recombinase and a RepA_N initiator, as well as erythromycin resistance, was found to be transduced by a phage at a higher frequency of transfer than could be achieved by filter mating [82]. This suggests that the boundary between viral satellite and conjugative element may not always be well defined, and individual elements may have some flexibility in their routes of mobilization. Since satellites typically do not encode their own structural genes, there is little to distinguish them from transposons or conjugative elements when one is making sequence based predictions. We anticipate that bacteriophage satellites will be found to be far more common than currently appreciated. Characterization of the PLE offers a window into these fascinating entities that shape the lives of their bacterial, and viral, hosts.

DATA AVAILABILITY

During review a direct link to access the data has been provided. The sequencing data from phage infected cells generated in this study have been deposited in the Sequence Read Archive database under bioproject accession code PRJNA577694.

ACKNOWLEDGMENT

We would like to thank members of Center for Structural Genomics of Infectious Diseases (https://csgid.org/) at the University of Toronto (Toronto, ON, Canada) and Northwestern University (Chicago, IL, USA) for assistance in protein crystal structure determination and Zdzislaw Wawrzak at Life Sciences Collaborative Access Team, Advanced Photon Source, Argonne National Laboratory, for X-ray diffraction data collection. The crystal structure determination in this project has been funded in whole or in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract Nos. HHSN272201200026C and HHSN272201700060C. We thank Brendan O'Hara, Stephanie Hays and Drew Dunham for the construction of several strains used in this study, as well as Stephanie Hays for preparing the sequencing libraries and helpful advice for data presentation. We thank members of the Seed lab for useful discussion and feedback, and the Komeili lab for use of their AKTA for protein purification. We would also like to thank the University of California, Berkeley QB3 Core Facility for assistance with whole genome sequencing. This work was supported by the National Institute of Allergy and Infectious Diseases [R01AI127652 to K.D.S.]; K.D.S. is a Chan Zuckerberg Biohub Investigator and holds an Investigators in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Fund. Funding for open access charge: University of California, Berkeley Startup Funds.
Conflict of interest statement. K.D.S. is a scientific advisor for Nextbiotics, Inc.

Chapter 3

A family of viral satellites manipulates invading virus gene expression and affects cholera toxin mobilization [83]

Zachary K Barth**,** Zoe Netter, Angus Angermeyer, Pooja Bhardwaj, Kimberley D Seed

Summary

Many viruses possess temporally unfolding gene expression patterns aimed at subverting host defenses, commandeering host metabolism, and ultimately producing a large number of progeny virions. High throughput -omics tools, such as RNA-seq, have dramatically enhanced resolution of expression patterns during infection. Less studied have been viral satellites, mobile genomes that parasitize viruses and have far reaching effects on host-cell fitness. By performing RNA-seq on infection time courses, we have obtained the first time-resolved transcriptomes for bacteriophage satellites during lytic infection. Specifically, we have acquired transcriptomes for the lytic *Vibrio cholerae* phage ICP1 and all five known variants of ICP1's parasite, the Phage Inducible Chromosomal Island-Like Elements (PLEs). PLEs rely on ICP1 for both DNA replication and mobilization, and abolish production of ICP1 progeny in infected cells. We investigated PLEs impact on ICP1 gene expression and found that PLEs did not broadly restrict or reduce ICP1 gene expression. A major exception occurred in ICP1's capsid morphogenesis operon, which was downregulated by each of the PLE variants. This transcriptional manipulation, conserved among PLEs, has also evolved independently in at least one other phage satellite, suggesting that viral satellites may be under strong selective pressure to reduce the capsid expression of their larger host viruses. Surprisingly, PLEs were also found to alter the gene expression of CTXϕ, the integrative phage that encodes cholera toxin and is necessary for virulence of toxigenic *V. cholerae*. One PLE, PLE1, upregulated CTXϕ genes involved in replication and integration, and boosted CTXϕ mobility following induction of the SOS response. Our data show that PLEs exhibit conserved manipulation of their host-phage's gene expression, but divergent effects on CTXϕ, revealing that PLEs can influence both their hosts' resistance to phage and the mobility of virulence encoding elements.

Introduction

Viruses are selfish genetic elements that reprogram their host cells for viral reproduction. Turning host cells into viral factories requires viruses to implement both their own tightly regulated gene expression programs and manipulations of host gene expression. Viral genomes can vary from just a couple genes [84] to sizes rivaling those of cellular life [85], so the gene expression strategies of viruses are highly varied. Viral life cycles exist on a continuum of agency. Some, like the cholera toxin phage (CTXϕ), are relatively passive. CTXϕ exists as an integrated prophage within toxigenic *Vibrio cholerae*. CTXϕ is largely regulated by host stress and virulence regulons, producing cholera toxin during *V. cholerae* infection of mammalian hosts, and replicating during the *V. cholerae* SOS response to DNA damage [86,87]. Aside from coding the two cholera toxin subunits, CTXϕ possesses a minimalist genome with just seven additional genes, five of which are structural or involved in virion morphogenesis. Upon induction, CTXϕ initiates its replication off the host chromosome. Assembled particles are released through host secretion machinery without killing the cell, allowing horizontal and vertical CTXϕ propagation [87]. In contrast, lytic phages are lethal to their hosts and may shut down host gene expression to maximize expression of their own genes [88]. These mechanisms often unfold in a concerted and controlled manner to give rise to tight, temporal patterns of gene expression during infection; as has long been evidenced through targeted studies, and more recently through global analyses such as RNA-seq [89–95].

Less explored are the transcriptional patterns of viral satellites. These subviral elements parasitize viruses in a similar way to how viruses parasitize their host cells. Like viruses, viral satellites are found in all domains of life and impact their hosts in profound ways. Viral satellites can partially or completely abrogate virion production by the viruses they parasitize [29,37,44], and can reduce or worsen disease in multicellular organisms [34,35]. Unicellular organisms can be protected against viruses on the population level by endogenous viral satellites, but the efficacy of protection varies depending on the specific virus and satellite genotypes, and infection context [29,38,43,96]. Given their broad distribution and importance for both their cellular and viral hosts, it is desirable to decipher how the reproductive programs of viral satellites intersect with and differ from the programs of the viruses they parasitize.

A prime model for mechanistic and evolutionary insights into viral satellites are the Phage Inducible Chromosomal Island-Like Elements (PLEs) found in toxigenic *V. cholerae*. PLEs parasitize ICP1 [29], a lytic myophage that is the predominant phage in cholera patient stool samples [26]. Following ICP1 infection, PLEs excise from the host chromosome, and replicate to high copy [29]. Successful PLE parasitism does not abrogate cell lysis, but results in the complete restriction of ICP1 and the release of PLE transducing particles [29] (Supplemental Fig 2.1). The tractable genetics of *V. cholerae* facilitates mechanistic studies of PLE gene products, and insights have been gained regarding chromosomal excision of PLE [55], PLE DNA replication [30,78], and PLE manipulation of lysis kinetics during infection [97]. Notably, ICP1 genome editing is also possible [98], allowing manipulation of both sides of this host-parasite relationship.

To date, five distinct PLEs have been identified within the genomes of *V. cholerae* isolates recovered from cholera patient stool samples dating back to the 1940s [29]. Each individual PLE occurs in isolation; no *V. cholerae* isolate has been found to harbor more than one PLE, and PLEs typically dominate for a time before disappearing and being succeeded by a new PLE genotype [29]. Four of the five known PLEs are integrated into repeats of the superintegron, an array of selfish and mobile elements in the *V. cholerae* small chromosome [29]. PLEs mobility, along with the extensive and growing library of ICP1 isolates, allows PLEs to be compared in shared strain backgrounds during infection by contemporaneous and non-contemporaneous ICP1 isolates [29,55,78]. These experiments have shown that PLE and ICP1 are engaged in a co-evolutionary arms race, with different pairings of PLEs and ICP1 isolates having different infection outcomes. The sole understood method through which ICP1 can overcome PLEs is the ICP1 encoded CRISPR-Cas system, and deletion of that system broadens the PLE and ICP1 interactions that can be studied [29,38,56]. Pairing PLEs against the same host-virus in an isogenic host-cell background allows us to probe for convergence and divergence in how PLEs exploit and restrict ICP1. Thus, the ICP1-PLE system is a powerful model for exploring co-adaptations between a virus and its satellite and is unparalleled for tracking how these adaptations have shaped the evolution of these warring elements.

So far, few insights have been gained into the gene expression programs of ICP1 and PLEs. PLE1 expresses its integrase in uninfected cells [55], expression of PLE1's replication initiator, RepA, is induced following infection of ICP1 [30], and the PLE's lysis modulator, LidI, is detectable by Western blot late during ICP1 infection [97]. The PLE integrase's recombination directionality factor (necessary for directing integrase excision activity) is PexA, an ICP1 protein whose native function is unknown, but whose expression can be detected by 5 minutes post infection [55]. While these limited observations have provided insight into key PLE and ICP1 genes, the gross expression patterns of ICP1 and PLEs remain unknown. Until now, we have not known the degree to which ICP1 alters cellular expression patterns, and whether PLEs alter ICP1's gene expression or reproduce and restrict ICP1 without such alterations. To address these questions, we performed RNA-seq on *V. cholerae* infected by ICP1 over the course of the infection cycle. We sequenced the transcriptome of ICP1 infection in the absence of PLEs, as well as in the presence of each of the five PLEs. This work deciphers ICP1's transcriptional program, the transcriptional program of each PLE, and the *V. cholerae* host transcriptome in each infection context. To our knowledge, this is the first detailed analysis of a viral satellite transcriptome during infection. Following ICP1 infection PLEs exhibit remarkable conservation of temporal transcription patterns and targeted alteration of ICP1 transcription. The patterns described here suggest that like many viruses, viral satellites like PLE have evolved to carefully coordinate gene expression. In contrast, when we compared uninfected PLE strains, we observed disparate interactions between PLEs and other mobile genetic elements in the *V. cholerae* genome. Surprisingly, most PLEs increase expression of the CTXϕ repressor *rstR*; however, the most recently circulating PLE, PLE1, upregulates CTXϕ's replication and integration factors, which we show enhances the mobility of CTXϕ. Collectively our

findings show that successive PLEs have conserved interactions with ICP1 and divergent interactions with CTXϕ, providing insights into how satellites manipulate the gene expression of their host viruses, and how they shape the evolution of their host cells.

Materials and Methods

Strains and culture conditions

All *V. cholerae* strains, including PLE(+) variants, used in this study are derived from E7946 to ensure comparisons in an otherwise isogenic background. Bacteria were grown on LB agar plates and in LB broth with aeration at 37°C. A detailed list of all strains used throughout this study can be found in Supplementary Table 2.1. ICP1_2006_E engineered to lack CRISPR-Cas (ΔCRISPR, *Δcas2-3*) [55] was used for all experiments. Phage titers were determined using a soft agar overlay method wherein ICP1 was allowed to adsorb to *V. cholerae* for 10 min at room temperature before the mixture was added to molten LB soft agar (0.3%) and poured onto 100 mm × 15 mm LB agar plates. Plaques were counted after overnight incubation at 37°C.

Generation of mutant strains and constructs

V. cholerae mutants were generated through natural transformation as described previously [99]. For antibiotic marked gene knockouts and overexpression constructs, splicing by overlap extension (SOE) PCR was used.

Sample collection for RNA-seq

Strains were grown to stationary phase in 2mL cultures before being back diluted to an OD_{600} of 0.05 in 6mL LB broth. Strains were then grown to an OD_{600} of 0.47 in 16x 150mm culture tubes with a Biochrom® Ultrospec 10 (equivalent to OD_{600} of 0.3 with a 1cm path length) before initial sample collection and phage infection at a multiplicity of infection (MOI) of 5. Immediately prior to infection, and then at 4, 8, 12, and 16 minutes post infection, 1mL of culture was taken and mixed with 1mL of ice cold methanol, before returning the remaining culture to the incubator. The sample and methanol mixtures were pelleted at 21,694 xg at 4°C for 2 minutes, aspirated, washed with 1mL ice cold 1X phosphate buffered saline (PBS), and then pelleted and aspirated again. Pellets were snap frozen in liquid N_2 and stored at -80°C until RNA-isolation.

RNA-isolation

RNA was extracted from samples using the Purelink™ RNA Mini Kit (Thermo-Fisher) and DNA was removed from isolated RNA samples using the TURBO DNA-*Free*™ Kit (Thermo-Fisher).

cDNA Library generation and sequencing

RNA samples were submitted to the University of California Berkeley QB3 Core facility for cDNA library generation and sequencing. Ribosomal DNA was removed with an Illumina Ribo-Zero rRNA Removal Kit (Bacteria) prior to cDNA generation. An S220 Focused-Ultrasonicator (Covaris®) was used to fragment the DNA, and library preparation was performed using the KAPA Hyper Prep kit for DNA (KK8504). Truncated universal stub adapters were used for ligation, and indexed primers were used during PCR amplification to complete the adapters and to enrich the libraries for adapter-ligated fragments. Samples were checked for quality on an AATI (now Agilent) Fragment Analyzer. Samples were then transferred to the Vincent J. Coates Genomics Sequencing Laboratory, another QB3-Berkeley Core Research Facility at UC Berkeley, where Illumina sequencing library molarity was measured with quantitative PCR with the Kapa Biosystems Illumina Quant qPCR Kits on a BioRad CFX Connect thermal cycler. Libraries were then pooled evenly by molarity and sequenced on an Illumina HiSeq4000 150PE flowcell. Raw sequencing data was converted into fastq format, sample specific files using the Illumina bcl2fastq2 software on the sequencing centers local linux server system.

RNA-seq analysis

For each sample library, sequencing reads were mapped to separate *V. cholerae*, ICP1_2006E (ΔCRISPR, Δ*cas2-3*), reference files, as well as files for the appropriate PLE genome in CLC Genomics Workbench version 12. Default RNA-seq mapping settings were used, with the exception that multiple mapping of individual reads was disabled. As noted previously [92,95], RNA-seq of lytic infections possess specific challenges because there are multiple genomes (two in most cases, three in the presence of a viral satellite such as PLE), undergoing changes in both their share of total transcripts in culture, as well as the relative expression of their genes. To address this, gene expression was normalized on a per genome basis for differential expression analysis. Differential expression analysis was performed using the DESeq2 [100] R/Bioconductor package with default parameters. For data visualization heat maps of log2 TPM values were plotted using the aheatmap function from the NMF R package. Volcano plots were generated using the EnhancedVolcano package and function. Our reads counts and DESeq2 results are provided within our supporting information (S1 Data). The accession numbers for reference sequences used for mapping can be also be found in our supporting information (Supplementary Table 1.5).

For the generation of reads tracks, RNA Seq reads were mapped to the reference sequences using bowtie2 v2.3.4.1 [58], with the following settings: "--end-to-end --very-sensitive --no-unal --no-mixed --no-discordant". For each sample, read coverage was normalized to sequencing depth and replicates were then averaged. Sequence data for samples used in this work can be found in the Sequence Read Archive under the Bioproject ID: PRJNA609114.

CTX transduction assays

V. cholerae PLE(-), PLE1, and PLE2 CTX(+) donor strains were modified by replacing *ctxAB* with a kanamycin resistance cassette. CTXφ production was induced in these strains by growing up to OD_{600}=0.3 followed by 16 hour incubation at 37°C with aeration in LB supplemented with mitomycin C (20ng/mL) (Sigma). *V. cholerae* CTX(-) recipient strains were engineered to harbor a cassette inserted in the *lacZ* locus containing a spectinomycin resistance gene and *toxT* under control of P_{tac} and a theophylline-inducible riboswitch. Recipient strains were grown to OD_{600}=0.3, then induced with addition of 1.5mM theophylline and 1mM IPTG at 37°C for 16 hours in LB+10mM $MgCl_2$ without agitation. After mitomycin C treatment, donor strains were centrifuged for 3 minutes at 5,000 x *g* twice to ensure maximum removal of donor cells from CTXφ-containing supernatant. Cleared donor supernatants were mixed with recipient cultures 1:4 and incubated at 37°C for 1 hour without agitation. Transduction mixtures were plated on LB plates supplemented with kanamycin (75µg/mL) and spectinomycin (100µg/mL) and incubated overnight at 37°C to quantify transductants.

Western blots

Isogenic *V. cholerae* strains either lacking PLE or with an integrated PLE (1-5) were grown to an OD_{600}=0.3 and infected with ICP1_2006E_△CRISPR *Δcas2-3* at MOI=1 and returned to the incubator at 37°C with aeration. At 16 minutes post phage addition, 1mL of infected culture was collected and mixed with an equal volume of ice-cold methanol. Samples were centrifuged at 5,000 x *g* for 10 minutes at 4°C to pellet infected cells. Pellets were washed once with ice-cold PBS and resuspended in lysis buffer (50mM Tris, 150mM NaCl, 1mM EDTA, 0.5% Triton X-100, 1x protease inhibitor (Thermo Pierce™ Protease and Phosphatase inhibitor tablet)). Protein concentration was quantified with Pierce BCA Protein Assay Kit (Thermo). 30µg of total protein sample was mixed with Laemmli buffer (Bio-rad) and boiled at 99°C for 10 minutes. Samples were run on Any-kD TGX-SDS-PAGE gels (Bio-rad) and transferred onto nitrocellulose membranes with Transblot Turbo Transfer system (Bio-rad). Custom primary peptide antibody generated in rabbits against ICP1 capsid (Gp122, YP_004251064.1) (GenScript) was diluted 1:1500. Band detection was conducted with a goat α-rabbit-HRP secondary antibody (Bio-rad) at 1:10,000 followed by development with Clarity Western ECL Substrate (Bio-rad) and imaging on a Chemidoc XRS Imaging System (Bio-rad).

Results and Discussion

V. cholerae's response to ICP1 infection

The ICP1 infection cycle takes approximately 20 minutes to produce a burst of nearly 90 infectious virions [29]. To capture the temporal range of ICP1's infection cycle, we took samples for RNA sequencing immediately prior to infection and 4, 8, 12, and 16 minutes post infection. Producing a large number of virions in a short period of time would

presumably require substantial changes to the host cell transcriptome, and we see such changes occur during ICP1 infection. At 4 minutes post infection there are already dramatic changes to *V. cholerae's* transcriptome, with 17.2% (658/3827) of genes differentially regulated compared to uninfected cells ($Q \leq 0.1$) (Fig 2.1A and Supplemental Table 2.2). At this 4 minute timepoint, slightly more host transcripts are predicted to be upregulated (345) than downregulated (313). When *V. cholerae* gene expression across infection is normalized to transcripts per kilobase million (TPM) and visualized by heatmap, it appears that the bulk of *V. cholerae* genes are decreasing in transcript abundance following infection, and a small subset is being upregulated (Fig 2.1B). We interpret the difference between our significant differential expression analysis results and TPM normalized expression profile to result from the differential expression analysis assuming a negative binomial distribution for gene expression changes [100]. While such a model is appropriate for most RNA-seq applications where the majority of genes are not differentially expressed, it may underreport the extreme transcriptional changes that can occur during lytic viral infection. These extreme changes are reflected by the changes in reads abundance over the course of infection. By 4 minutes post infection, ICP1 contributes to more than a quarter of total RNA reads within the culture, and by 16 minutes post infection, ICP1 reads comprise more than 80% of total reads in the culture (Fig 2.1C). These changes appear even more extreme when the relative size of the *V. cholerae* and ICP1 genomes are taken into account (Fig 2.1D). Taken together, we interpret this data to show that ICP1 affects *V. cholera*e transcription by globally reducing *V. cholerae* gene expression while upregulating the expression of a subset of genes.

V. cholerae's most downregulated genes are enriched for tRNA and rRNA processing genes, while several different gene groups are enriched among upregulated transcripts (Fig 2.1A and Supplemental Table 2.2). The most dramatic differential expression of *V. cholerae* genes occurred in the ArgR regulon responsible for arginine biosynthesis. Arginine biosynthesis genes were highly upregulated upon ICP1 infection, with *argB, argC,* and *argF* expression increasing more than 100-fold at 4 minutes post infection (Fig 2.1A and Supplemental Table 2.2). Similarly, the arginine transport genes were the most highly upregulated genes encoded in the *V. cholerae* small chromosome. The Na^+/H^+ antiporter *nhaA* was also strongly upregulated. Increases were also seen for genes relating to other amino acid biosynthesis and transport, sulfur compound metabolism, ATP production, flagellar synthesis and motility, and cell division (Supplemental Table 2.2). In contrast to the decrease in tRNA and rRNA processing genes, we see an increase in ribosomal protein coding genes and other genes involved in translation (Fig 2.1A and Supplemental Table 2.2). Several functional gene classes had many members upregulated and downregulated, for example a large number of genes involved in transport were both up and downregulated, likely reflecting the difference in metabolic needs for lytic virus production versus normal cell growth. We also saw differential expression of cell envelope genes, with a decrease in some lipopolysaccharide biosynthesis genes and an increase in the mannose-sensitive hemagglutinin (MSHA) pilus associated with estuarine growth (Figure 2.1A and Supplemental Table 2.2) [101,102].

It is hard to predict with certainty which transcriptional changes are a defensive host response to infection and which are due to transcriptional manipulation on the part of ICP1. Many of the differentially expressed genes relate to phenotypes regulated by cyclic di-GMP (di-cGMP) [103]. Recently, another cyclic dinucleotide, cyclic GMP-AMP (cGAMP), which was first discovered in *V. cholerae* [104], has been linked to phage defense [105]. It is interesting to consider that ICP1 infection may be influencing cyclic di-nucleotide signaling by triggering a host defense system, though in this case the defense is not successful. Regardless of the source of these alterations, the expression changes we see are not fully consistent with a shift toward high or low levels of either cyclic dinucleotide. We see several gene changes consistent with high di-cGMP [103] (increased MSHA biosynthesis, increased cold shock, decreased heat shock, and increased type VI secretion expression (Supplemental Table 2.2)) and some changes consistent with low di-cGMP [103] (increased flagellar synthesis and increased expression of the virulence regulator ToxT (Supplemental Table 2.2)). An increase in MSHA is also consistent with low cGAMP, but we also see downregulation of some chemotaxis associated genes, and chemotaxis is repressed by high cGAMP [104]. These changes suggest that the host cell may be receiving competing regulatory inputs that may act above or below the level of cyclic di-nucleotide signaling.

The strongly increased expression of arginine metabolism and transport genes is especially curious and could have positive or negative effects on phage production. ICP1 may be upregulating arginine metabolism to drive the production of purines or polyamines. Arginine often serves as the precursor for polyamine synthesis [106]. Purines may be a limiting resource for phage genome replication, while polyamines are found in the capsids of multiple phages, potentially aiding in DNA condensation, packaging, and ejection [107,108]. We do not see upregulation of genes specific for synthesis of the most common polyamine, putrescine, but the ornithine decarboxylase which is responsible for putrescine synthesis is regulated post-translationally in *Escherichia coli* and mammals to allow rapid adjustment of polyamine pools [106,109]. We also see upregulation of the *V. cholerae* inducible lysine decarboxylase (S2 Table), and in at least one bacterial system, a lysine decarboxylase is able to use ornithine as a substrate [110]. Alternatively, arginine has been found to have a deactivating effect on phage virions under certain conditions [111–113], so the upregulation of arginine may be an attempt by the host cell to curb phage production.

Establishing the ICP1 transcriptional program

Once we examined ICP1's effect on *V. cholerae* transcription, we next sought to document ICP1's transcriptional program. ICP1 has an approximately 126kb genome and more than 200 predicted open reading frames [52]. Less than a quarter of ICP1's putative coding sequences have activities or functions that can be predicted through bioinformatic analysis. As is common for viruses, ICP1's genes fall into distinct temporal groupings based on the timing of peak expression. The early genes, which we define as those that show peak expression at 4 minutes post infection, consist mostly of short genes averaging less than 330bp and encoding hypothetical proteins (Fig 2.2 and Supplemental Table 2.3). It is difficult to infer the function of these genes, but there are

several short immediate early genes in other phage systems that are known to have a role in host cell takeover [114,115]. The next grouping, middle early genes with peak expression at 8 minutes post infection, is mostly comprised of nucleotide metabolism genes including ICP1's DNA polymerase (Fig 2.2 and Supplemental Table 2.3). The 12 minute time point captures a transitional period in ICP1 transcription. High expression of the middle early genes continues, while expression of late genes has begun but not yet peaked (Fig 2.2). Few genes hit peak expression at 12 minutes, though an exception are a subset of nucleotide metabolism genes (within the range *gp176–gp211*). Finally, late genes with peak expression at 16 minutes post infection are comprised primarily of putative structural genes and genes known to be involved in lysis (Fig 2.2 and Supplemental Table 2.3), [97]. The lysis, capsid, and tail genes occur in three separate clusters, all encoded on the (-) strand (Fig 2.2 and Supplemental Table 2.3). Previously, ICP1 rolling circle replication was predicted to proceed in the (-) direction based on late infection DNA coverage skews [30]. Since ICP1's late genes are transcribed off of the (-) strand, they may be transcribed off of the rolling circle replication template. Such an arrangement would be consistent with the preference for co-directional transcription and DNA leading strand synthesis observed in many bacterial systems [116], and thought to help preserve genome integrity by avoiding replication and transcription conflicts [117]. Overall, the expression patterns we see for ICP1 are consistent with what is known about lytic phage development in general and ICP1's life cycle in particular.

PLEs exhibit a conserved transcriptional program

Having established the transcriptional program of ICP1, we next sought to examine transcriptional patterns in PLEs. The five PLEs share a similar gene organization. Proximal to the PLE's left attachment site is a gene cluster (denoted here C_{int}) that includes the PLE integrase, *int*. Immediately downstream of this gene cluster is PLE's sole cluster of negative strand encoded genes, here called C_L for 'left' cluster (Fig 2.3A). This (-) sense cluster flanks an approximately 3kb non-coding region, the right most quarter of which contains the PLE origin of replication [30]. Flanking the PLE origin of replication is a putative *marR*-like gene, and this is followed by two additional rightward facing clusters (C_{R1} and C_{R2}). Exceptions to this arrangement occur with PLE1's standalone *int*, a transposase in PLE2's large non-coding region, and the absence of a *marR*-like gene in PLE3 [29].

Consistent with the PLEs providing interference against this ICP1 isolate [29], we find that all PLEs are transcriptionally activated following infection. Paralleling PLEs' organizational similarities, the transcriptional patterns of PLEs are highly conserved once each PLE has been activated. In uninfected samples, PLEs show some expression of *int*, and, if it is present, the *marR*-like gene, while expression across C_L, C_{R1}, and C_{R2} is variable and often uneven (Fig 2.3B and Supplemental Figs 2.2-2.5). Read counts for PLE genes remain low at 4 minutes post infection (Supplemental Data 2.1), and then at 8 minutes post infection transcriptional patterns start to emerge. For each PLE, C_{R1} is strongly expressed at 8 minutes post infection, followed by strong expression of C_L at 12 minutes, and high C_{R1} expression at 16 minutes along with the *marR*-like gene (Fig 2.3 and Supplemental Figs 2.2-2.5). It should be noted that all PLE

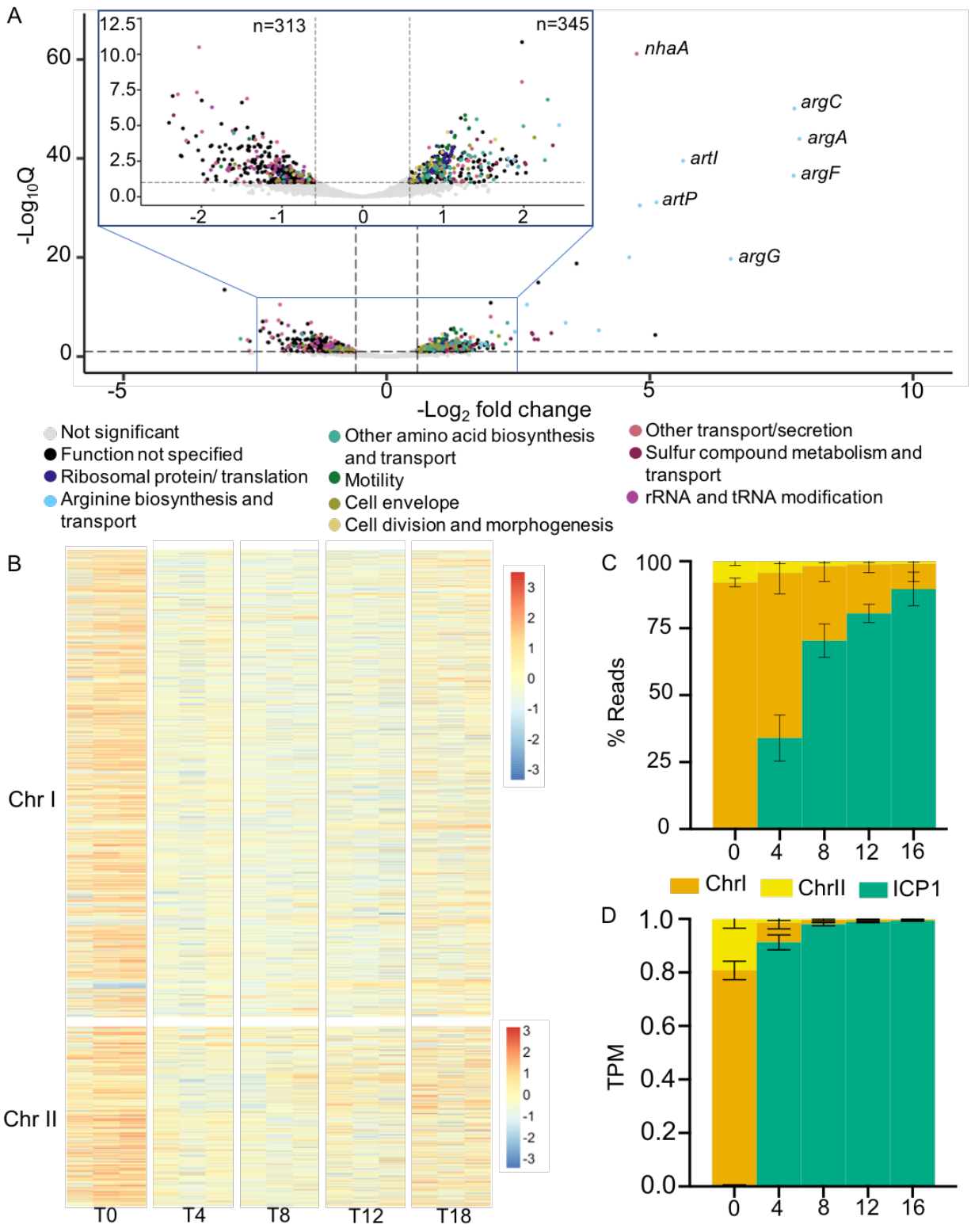

A
n=313
n=345
nhaA
argC
argA
artI
argF
artP
argG
-Log10Q
-Log2 fold change
Not significant
Function not specified
Ribosomal protein/ translation
Arginine biosynthesis and transport
Other amino acid biosynthesis and transport
Motility
Cell envelope
Cell division and morphogenesis
Other transport/secretion
Sulfur compound metabolism and transport
rRNA and tRNA modification
B
Chr I
Chr II
T0
T4
T8
T12
T18
C
% Reads
ChrI
ChrII
ICP1
D
TPM

Figure 2.1. The *V. cholerae* response to ICP1 infection. (*previous page*)
(A) Volcano plot showing gene expression changes between uninfected *V. cholerae* cultures, and the same cultures 4 minutes post infection. Full (bottom) and zoomed (top) views are provided to improve gene feature visibility. Genes are colored according to annotation. A significance cutoff of a Q value less than or equal to 0.1 and a $-\log_2$ fold change magnitude greater or equal to 0.585 (approximate to 1.5-fold) was used. **(B)** Heat map showing changes in *V. cholerae* gene feature $\log_2$ TPM over the course of ICP1 infection. Expression values obtained for three biological replicates are shown at each time point. ICP1 genes were excluded from TPM calculation to highlight relative changes in *V. cholerae* transcript abundance. Colors reflect the Z-score of each gene's $\log_2$ TPM value across replicates and time points. **(C)** Percent reads abundance for both *V. cholerae* chromosomes and ICP1 over the infection time course. **(D)** Reads normalized to a TPM value based on the total number of reads from each element, and the element's length. For all panels, results incorporate values obtained for three biological replicates.

transcripts continue to increase in abundance over the course of infection, though we see differences in the timing of peak gene transcription relative to other genes. This sustained global increase can likely be attributed to increased PLE copy number, as PLE replicates upwards of 1000-fold during ICP1 infection [29].

Early expression of C_{R1} is not surprising, given that this cluster contains PLE's *repA* gene which is necessary for PLE replication [30], and PLE1 replication was previously found to begin before 15 minutes post infection [29]. Interestingly, a highly conserved

C_{R1} gene, *orf12.1* in PLE1, has a different expression pattern than the rest of the cluster, peaking at 12 minutes post infection instead of 8. This suggests that *orf12.1* may be under different regulation than the rest of C_{R1}, and may be involved in the transition from early to late PLE gene expression. Overall, the conserved timing of expression of syntenic gene clusters across PLEs suggest that each cluster serves a distinct role in parasitizing ICP1. We suspect that the timing of PLE gene cluster expression has evolved to take advantage of ICP1's own transcriptional program. Coordination between PLE gene expression and ICP1's gene expression would be consistent with PLEs' reliance on ICP1 gene products for key steps of the PLE life cycle [30,55,78,97].

Non-coding RNAs are abundant in ICP1 and PLEs

Non-coding RNAs (ncRNAs) are a prominent feature of both ICP1 and PLE gene transcription. Surprisingly, the most abundant transcripts in the ICP1 and PLE transcriptomes are both predicted to be non-coding. The most abundantly expressed ICP1 transcript is encoded in an approximately 1kb orf-less space between *gp139*, the start of ICP1's lysis cluster [97], and *gp140* (Fig 2.4). The length of this transcript (~800bp based on the RNA-seq coverage), is comparable to the Giant- Ornate- Lake- and Lactobacialles-Derived (GOLLD) and the Rumen-Originating, Ornate, Large (ROOL) RNAs that have been found in many phage genomes [118]. These RNAs are frequently encoded near tRNAs, although this is not the case for ICP1 since it does not possess any tRNAs. The role of these large ncRNAs is unknown. In one *Lactobacillus brevis* prophage, a GOLLD RNA was found to accumulate during lytic infection, but was dispensable for phage production [118].

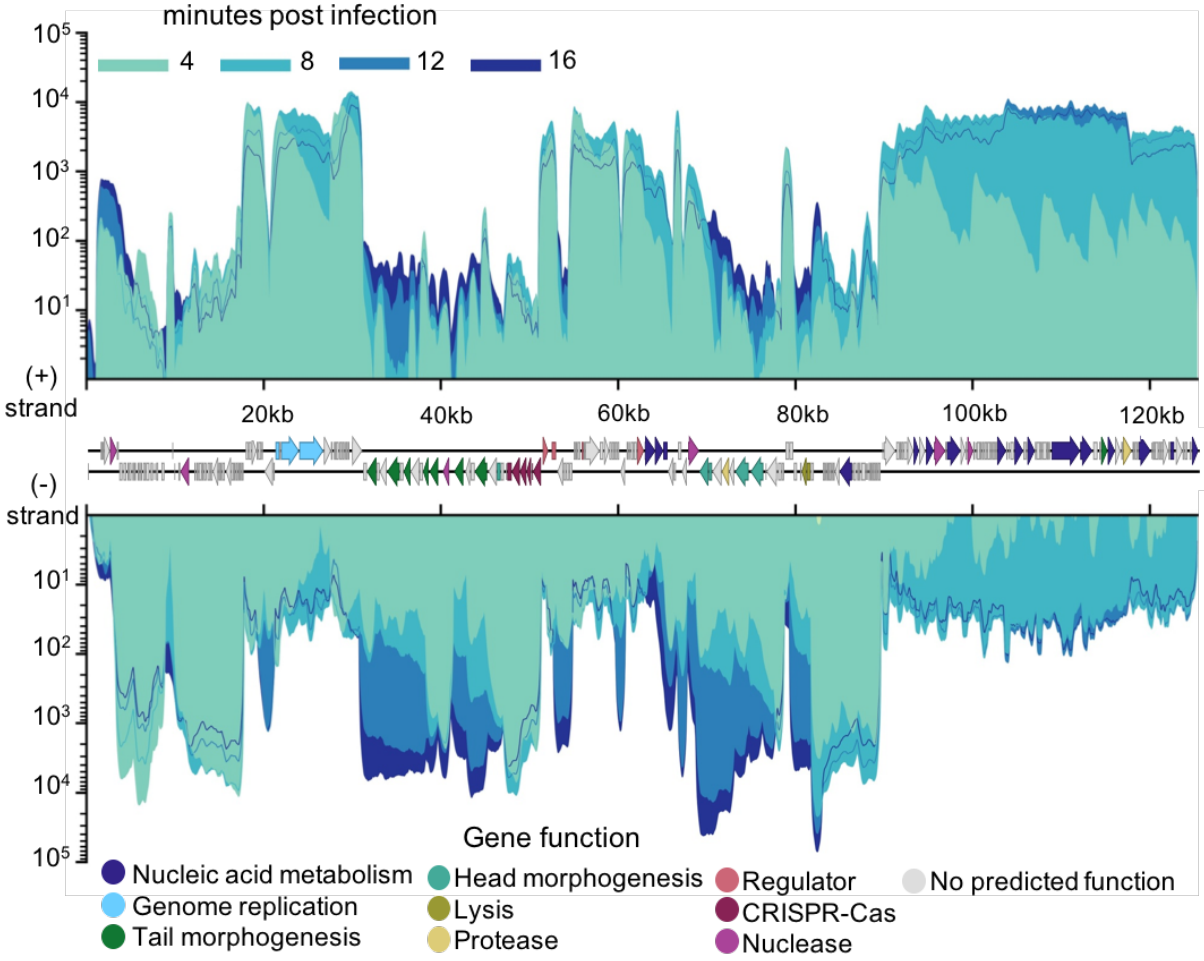


Figure 2.2. ICP1 transcriptional patterns.
ICP1's genome displaying average reads coverage over the course of infection of PLE (-) *V. cholerae*. Reads are color coded by time point. Gene features are colored based on known or putative gene functions. Results incorporate values obtained for three biological replicates.

In each PLE, the most abundant transcript is located between C_{R1} and C_{R2} (Fig 2.5 and Supplemental Figs 2.6-2.9). This transcript occurs between a set of inverted repeats, and for this reason, we have tentatively named the transcript the inter-inverted repeat (IIR) transcript (Supplemental Fig 2.10). Non-coding RNAs flanked by terminal inverted repeats often occur in miniature inverted-repeat transposable elements (MITEs), and these have been examined in several bacterial species [119]. This similarity suggests that the PLE IIR may have evolved from a selfish mobile ancestor. The PLE IIR transcript also has antisense homology to the leader sequences of several PLE ORFs (S10B Fig). Complementarity between a ncRNA and gene leader sequences is seen in the regulatory RNAs from phages P1, P7, and N15, as well as the phage satellites P4 and ϕR73 [120–122]. With the exception of N15's RNA which promotes the lytic cycle,

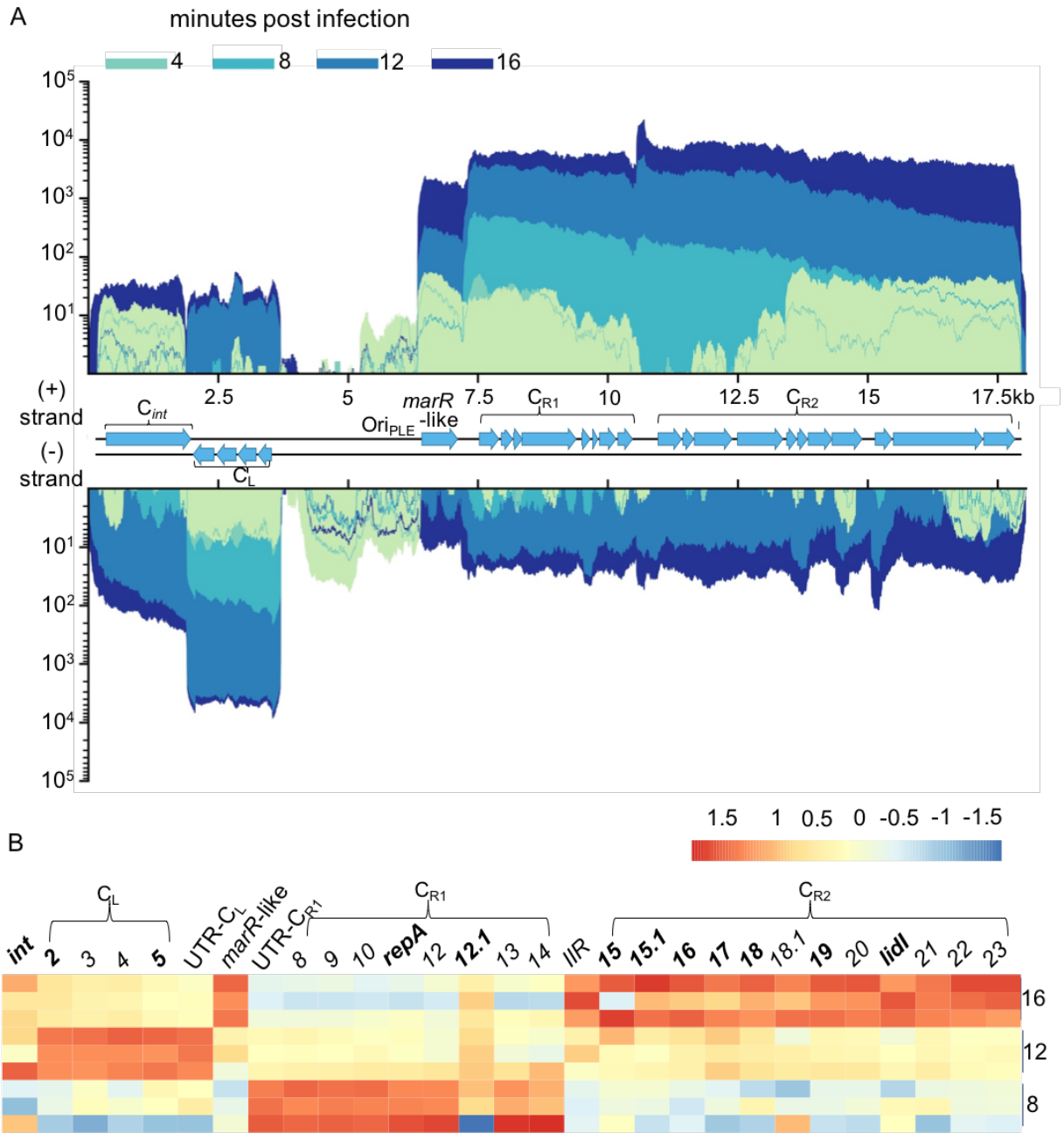


Figure 2.3. PLE transcriptome.
(**A**) PLE1's genome displaying average reads coverage over the course of infection. Reads are depicted on a logarithmic scale, to improve visibility of early expressed genes. Reads are color coded by time point. (**B**) Heat map of PLE1 gene expression over the course of infection. Color reflects the Z-score of each gene's log2 TPM value across replicates and time points. *V. cholerae* and ICP1 genes were excluded from TPM calculation to highlight relative changes in PLE1 transcript abundance. Core genes, protein coding genes with high conservation across PLEs, are bolded. Values for 8, 12, and 16 minutes post infection are shown. Results incorporate gene expression values obtained for three biological replicates.

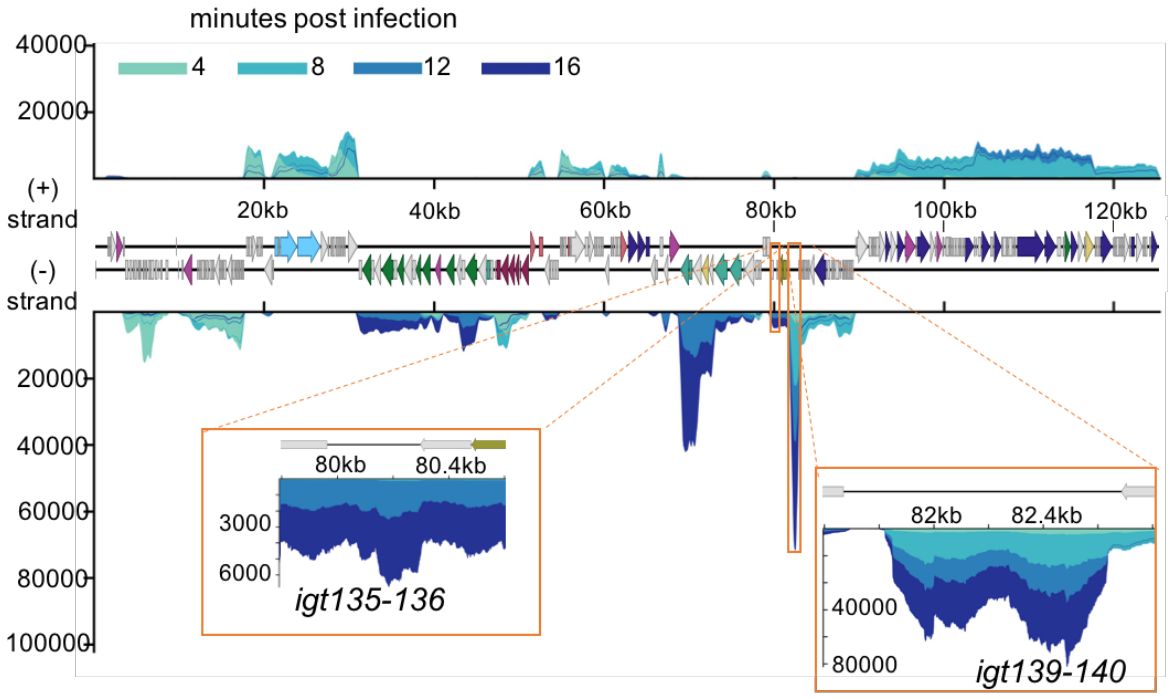


Figure 2.4. ICP1 noncoding RNA
ICP1's genome displaying average reads coverage over the course of infection on a linear scale. Reads are color coded by time point. ICP1 gene features are colored based on known or putative gene functions as in Fig 2.2. Regions boxed in orange show transcripts that lack predicted coding sequence. Results incorporate gene expression values obtained for three biological replicates.

these regulatory RNAs function to repress the lytic cycle. These roles appear unlikely in PLE, since there is an accumulation of the PLE IIR transcript as infection progresses, but it is not abundant until 12 minutes post infection, which is after PLE early gene expression (Fig 2.3B, Supplemental Figs 2.2.-2.5). The PLE IIR transcript's expression pattern appears more consistent with modifying gene expression during infection rather than acting as a global activator or repressor of PLE activity.

Additional transcripts without predicted coding capacity occur in both ICP1 and PLEs. An approximately 300bp region between *gp135* and *gp136* is transcribed in ICP1 during 12 and 16 minutes post infection (Fig 2.4). Within each PLE, we see abundant transcription approximately 200bp upstream of C_L and 150bp upstream of C_{R1} when these clusters are transcriptionally active (Fig 2.5 and Supplemental Figs 2.6-2.9). Though these 5' UTRs are not conserved on the sequence level, their occurrence in every PLE suggests conservation of function. The 5' UTR of transcripts is a common site for riboswitches [123] suggesting that these untranslated sequences may regulate expression of their downstream genes.

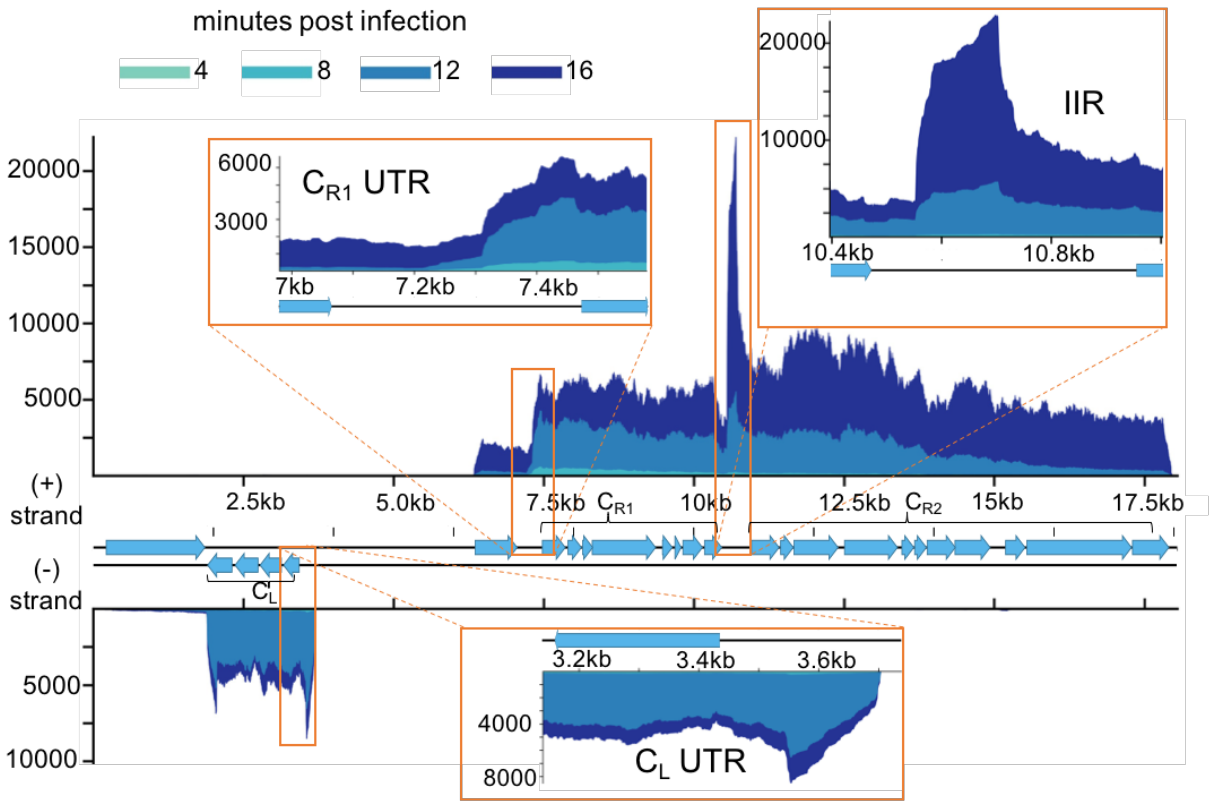


Figure 2.5 PLE1 noncoding RNA
PLE1's genome displaying average reads coverage over the course of infection on a linear scale. Reads are color coded by time point. Inserts depict detected transcripts that lack predicted coding sequence. Results incorporate gene expression values obtained for three biological replicates.

PLE-host interactions in the uninfected cell

By comparing transcriptomes of PLE(+) and PLE(-) strains prior to infection, we were able to determine whether PLEs affected transcription in their *V. cholerae* host prior to phage infection. All PLEs altered the transcription of genes neighboring the PLEs' integration sites (Fig 2.6A and Supplemental Fig 2.11). Additionally, PLEs altered the expression of several genes within the *V. cholerae* superintegron, including multiple toxin-antitoxin systems (Supplemental Table 2.4). Altered expression of superintegron encoded genes also occurred in the PLE2(+) strain, despite PLE2 being integrated outside of the superintegron [29]. These transcriptional changes may reflect cross talk between PLE-encoded genes or genes flanking the PLE integration site, and other genes encoded within the superintegron. In *V. cholerae*, both mobile genetic elements and genes in the superintegron are known to be repressed by the bacterial chromatin protein H-NS [124–126]. An alternative explanation for why we see changes in superintegron gene expression is that PLE integration alters the nucleoid architecture of this region through its own recruitment of H-NS. Notably, many of the superintegron genes are multi-copy, so the number of differentially regulated genes may be over-reported if the reads mapping cannot differentiate reads from multi-copy genes from different loci. This explanation seems particularly likely for scenarios where genes that

are distal to the PLE integration site are predicted to be differentially expressed, and have paralogs located proximal to the integration site. Such a pattern is seen with PLEs 3 and 5 (Supplemental Table 2.4), but can only explain a small number of the differences we see in the superintegron.

Rather surprisingly, we also found that PLEs alter the expression of genes encoded by integrative mobile elements exploiting Xer (IMEXs) (Supplemental Fig 2.11). IMEXs are mobile elements that utilize host Xer recombinases to integrate into the chromosome dimer resolution or *dif* sites of their host-cells, located near the chromosome replication terminus [127]. The strain of *V. cholerae* used in this study has three separate IMEX's: CTXϕ, CTXϕ's satellite RS1ϕ, and the toxin-linked cryptic element (TLC), integrated twice in tandem next to CTXϕ (Supplemental Fig 2.12). RS1ϕ is largely redundant with sequence within CTXϕ. RS1ϕ and CTXϕ have their own copies of the replication initiator *rstA,* a gene required for integration named *rstB*, and the CTXϕ master repressor *rstR* [86,87]. All these IMEXs are integrated on chromosome I, so the PLEs transcriptional effects on these elements must be acting *in trans*. Relative to the PLE(-) strain, most PLEs increased expression of both the RS1ϕ and CTXϕ copies of *rstR*. These same strains also showed upregulation of the TLC gene *tlcR.* An exception to this pattern occurred in PLE1, where rather than *rstR* upregulation we observed upregulation of the CTXϕ replication genes *rstA* and *rstB* (Fig 2.6A). This observation prompted us to question whether upregulation of *rstA* and *rstB* could prime CTXϕ for mobilization following activation of *V. cholerae's* SOS response. To test this, we used an antibiotic marked copy of CTXϕ and found that following mitomycin C treatment, the presence of PLE1 increased the production of CTXϕ transducing units 10-fold relative to a strain with no PLE or a strain with PLE2. (Fig 2.6B). These results reveal potentially far reaching effects that mobile elements can have on each other, as well as the hosts they share. Beyond providing the host-cell population with immunity to ICP1 phages, PLEs' integration can enhance the mobility of other mobile genetic elements and, by extension, virulence genes. Thus, PLEs may affect *V. cholerae* fitness in ways that are distinct from their own anti-phage activity and relevant to cholera epidemiology.

PLEs selectively manipulate ICP1 transcription

Having detailed ICP1's and PLEs' transcriptional programs and PLEs' transcriptional effects in uninfected cells, we sought to evaluate whether PLE disrupted ICP1 gene expression. Remarkably, although PLEs abolish ICP1 production [29] we found that PLEs do not broadly restrict or alter ICP1 transcription. At 16 minutes post infection during maximum PLE expression, PLE transcripts comprise roughly 10% of the transcriptome while ICP1's proportion of reads still sits around 80% (Fig 2.7A). When normalized to the genome size, PLE1 TPM approaches parity with that of ICP1, while the other PLEs achieve a bit less (Fig 2.7B and Supplemental Figs 2.13-2.16) indicating that the relative transcriptional activity of PLEs do not exceed that of ICP1. This is in stark contrast to previously reported DNA levels at 16 minutes post infection, where the amount of PLE1 DNA exceeds that of ICP1, and overall ICP1 DNA replication is substantially reduced by PLE1 [30]. During ICP1 infection of PLE1 positive cells there is a loss of ICP1's rolling circle replication [30]. PLE1's disparate effects on ICP1

replication and transcription can be reconciled by the model that ICP1 late genes are transcribed off the same strand of DNA that serves as the template for rolling circle replication. A block in ICP1 rolling circle replication would not impede ICP1 transcription if the newly synthesized DNA is not expressed, but would permit PLE to interfere with ICP1 packaging as has been hypothesized [30].

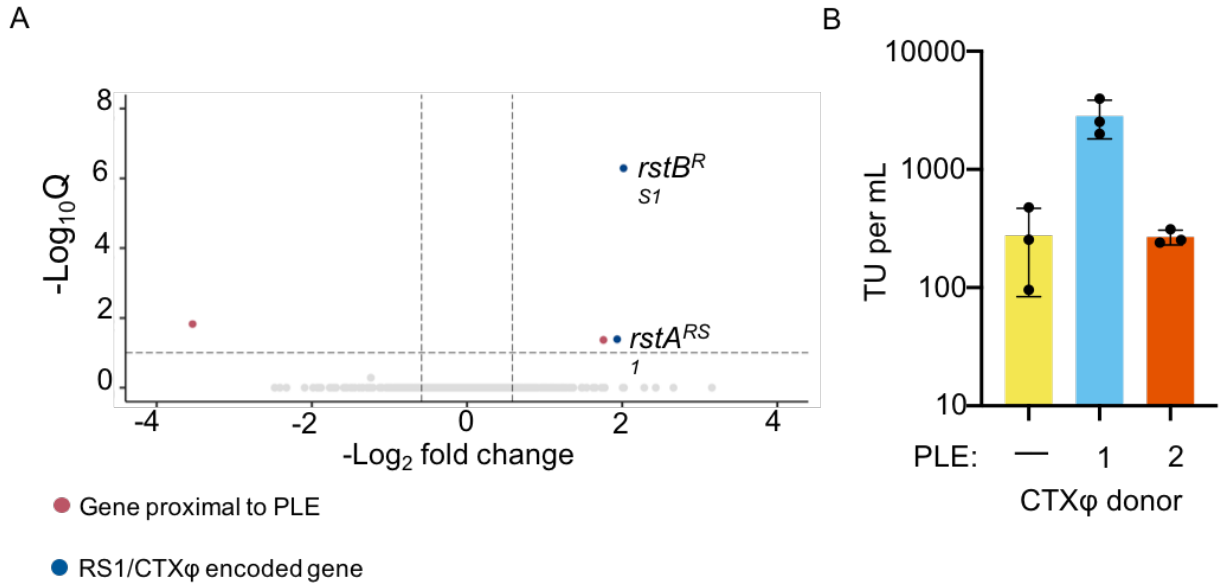


Figure 2.6 CTX production is upregulated in PLE1 positive strains
A. Volcano plot showing differential regulation of *V. cholerae* genes in uninfected PLE1(+) cultures relative to PLE (-) cultures. Genes within 2kb of the PLE integration site are colored red, and genes encoded in RS1 and CTXφ are colored blue. Genes that are not significantly differentially regulated are colored light gray. A cutoff of a Q value less than or equal to 0.1 and a -log2 fold change magnitude greater or equal to 0.585 (approximate to 1.5-fold) was used. **B.** The transduction units per mL of CTXφ produced by PLE(-), PLE1(+), and PLE2(+) cultures following induction by mitomycin C.

In addition to potential mechanistic explanations for how PLE is able to substantially restrict ICP1 replication without broad disruptions of ICP1 transcription, it is interesting to consider why such a discrepancy in PLE activity would be favorable. PLEs are not very transcriptionally active until 8 minutes post infection (Fig 2.3A, Supplemental Figs 2.2-2.5, and Supplemental Data 2.1). From 8 minutes post infection onward, ICP1 transcripts are comprised largely of nucleotide metabolism genes, genes involved in DNA replication, and genes encoding virion structural components (Fig 2.2). Aside from RepA, the replication initiation factor that directly interacts with the PLE origin of replication, PLE does not appear to encode dedicated replication machinery [30]. Further, PLE has been shown to rely on at least some ICP1 gene products for replication [30,78]. PLE also does not encode identifiable structural genes, and requires the same viral receptor as ICP1 for mobilization [29], suggesting that like other phage satellites [32], PLE is packaged within the same virion structural components as its host-phage. The reliance on ICP1's virion production machinery incentivizes PLE to

allow ICP1's transcriptional program to progress relatively unperturbed, since ICP1 is already producing the infrastructure for robust virion production. Our data suggests that rather than suppressing the production of ICP1 machinery and replacing it with PLEs' own, PLEs efficiently parasitize that machinery, redirecting it to PLE's own genome and somehow excluding ICP1. Such a strategy would allow PLE to benefit from ICP1's own reproductive adaptations, while restricting ICP1 propagation among the host-cell population.

While PLEs may benefit from permitting ICP1 gene expression, they also might benefit from re-tuning certain aspects of ICP1's transcriptional program to better fit their needs. Consistent with this hypothesis, we observed that one set of ICP1 genes has markedly decreased expression in the presence of all PLEs: *gp122-gp126* (Fig 2.7C). This group of genes is predicted to be responsible for ICP1 capsid morphogenesis. The genes *gp122* and *gp123* are predicted to encode ICP1's major capsid protein and a capsid decoration protein, respectively, and *gp125* is a predicted protease, likely providing the proteolysis necessary for procapsid maturation as occurs in most tailed phages [128]. Reduced expression of this gene cluster is seen in all PLEs, though the Q values we obtained from the differential analysis for PLE2 and PLE4 were less robust (Supplemental Figs 2.13-2.16). Nevertheless, when the production of ICP1's major capsid protein was assessed in the presence of each PLE via Western blot, we found that consistent with the differential transcription observed, all PLEs reduce ICP1 capsid production between 2 and 3-fold (Fig 2.7E and Supplementary Fig 2.17). These results show that specific reduction in capsid production is a conserved activity among PLEs.

The PLEs' downregulation of ICP1 capsid genes may be evolutionarily tied to remodeling of the virion capsid. Capsid remodeling, more specifically reduction in capsid dimensions, is a well-studied feature of phage satellite parasitism, having arisen in the phage satellite P4, the Gram-positive Phage Inducible Chromosomal Islands (PICIs), and the Gram-negative PICI's [32,42], though it has yet to be shown for PLEs. Because capsid remodeling has been observed in the three other lineages of tailed phage satellites, capsid remodeling may be a general feature of phage satellite biology, and is likely to be induced by PLEs. Capsid remodeling can restrict host-viruses by assembling virions that are too small for the full-sized genome. Additionally, capsid remodeling likely increases the horizontal mobility of satellites. PLEs are not predicted to encode their own large terminase, suggesting that they rely on ICP1 packaging machinery in addition to structural components. Given that ICP1's genome is about 7 times the length of PLE, and ICP1 is predicted to package virions in a head-full fashion [30], packaging PLEs into ICP1 sized capsids would result in an overabundance of PLE genomes in a single transducing particle. Excess genome packaging would reduce the number of transducing particles that could be assembled for a set level of PLE genome replication.

While reducing capsid size would benefit PLE by reducing the amount of excess PLE DNA packaged per PLE transducing particle, further benefits could be gained by tuning gene expression to reduce production of excess capsid. Prior analyses have shown that protein translation is expected to be the most energy intensive process for viruses within

PLE and ICP1's genome size range [129,130], and the major capsid transcript is by far the most abundant mRNA produced by ICP1 (Fig 2.2). Reducing the amount of capsid produced could help PLEs' recoup the costs of PLE gene expression and additional DNA replication which do not occur during ICP1 infections in the PLE(-) background.

Much as capsid remodeling is a recognized occurrence among tailed phage satellites, our results reveal an emerging pattern of satellite elements tuning down expression of phage late genes. Downregulation of capsid, at least relative to other structural genes, has arisen independently among multiple phage satellites. Satellite P4 encodes a gene named *psu* for polarity suppression unit. Psu acts as a coat decoration protein for the P2 capsid when remodeled by P4, increasing the stability of remodeled capsids [39]. Additionally, Psu is well characterized as a Rho-binding antitermination factor [131–133]. P2 regulates its structural genes through transcriptional attenuation: transcription frequently terminates before reading through the entire operon, so the genes near the front of their operon, including the capsid and scaffold, are expressed to a greater extent than downstream genes [134]. By preventing Rho-dependent termination, P4 retunes structural gene expression so that the ratio of capsid and scaffold relative to other structural genes is reduced, with the ratio of capsid and scaffold to terminase being 5 to 10-fold lower in the presence of P4 [135]. Initially, it was proposed that P4's tuning of structural gene transcription was the mechanism through which P2 capsids were remodeled, but later, remodeling was found to actually be caused by the Sid capsid scaffold, encoded in the same three-gene operon as Psu [136]. Though Psu has undergone extensive biochemical characterization, little has been uncovered about the evolutionary importance of P4 induced antitermination since it was found to be dispensable for capsid remodeling.

SaPIs provide an additional example of phage satellites repressing host-phage genes. In the SaPI host-phage 80α and the related phage 80, late genes are organized into a putative operon starting with the small terminase encoding gene *terS*, followed by additional packaging genes, then head morphogenesis genes, tail morphogenesis genes, and finally lysis genes. Some SaPIs have been found to repress *terS*, and it was inferred that there was also repression of the late gene operon [137] . It was noted that repression must be incomplete, since SaPIs rely on phage structural, lysis, and large terminase genes for their propagation. While repression of *terS* could benefit SaPIs by preventing the packaging of host-phage genomes, complete repression of the late operon would block SaPI particle production. Notably, only *terS* expression was measured for this operon. If internal promoters exist in this late operon, such that *terS* is silenced by SaPIs, head morphogenesis genes are repressed to an intermediate degree, and tail and lysis genes are unaffected, it would be consistent with the pattern of head morphogenesis repression seen in P4 and PLE, as well as the reproductive needs of SaPIs.

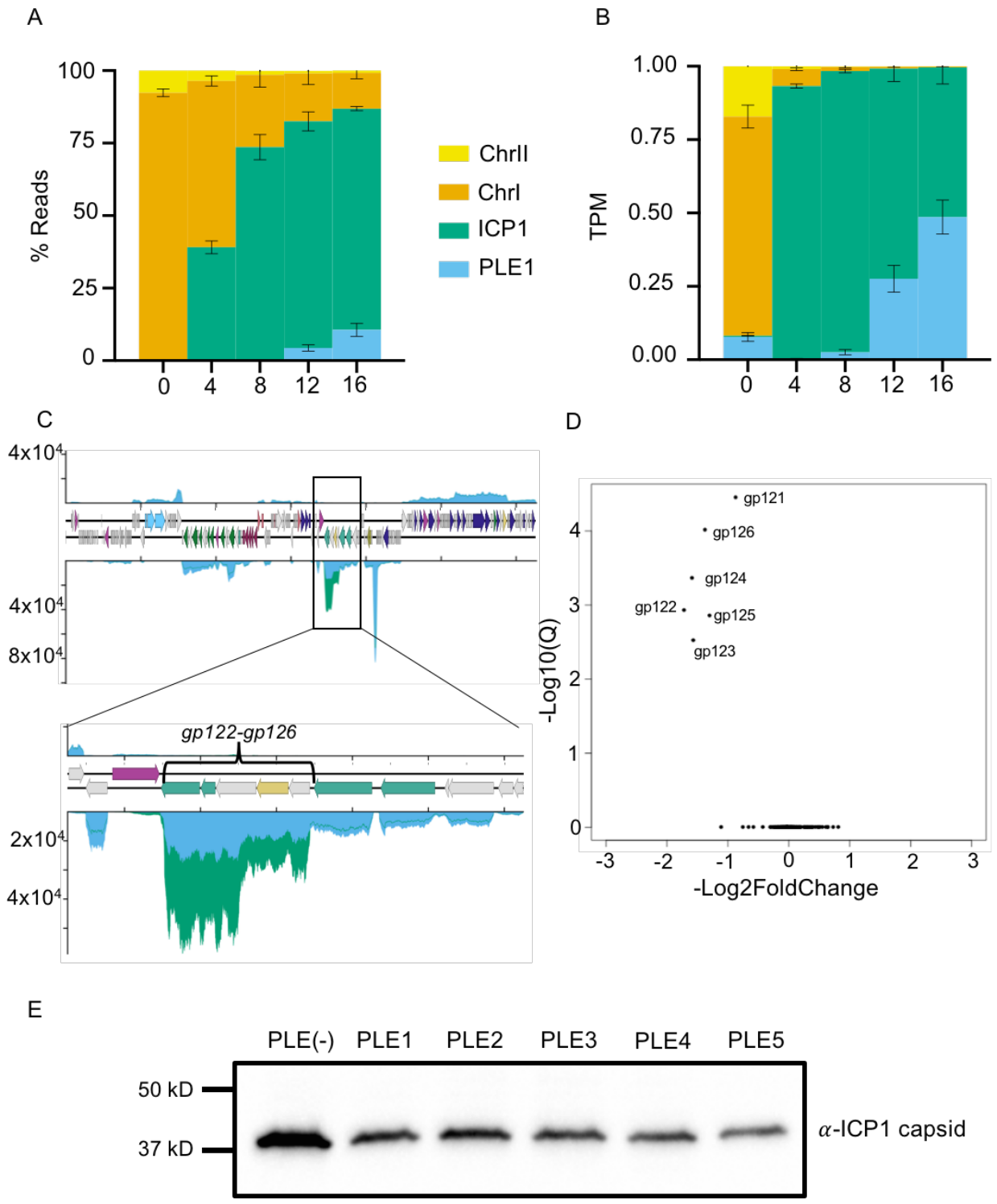


Figure 2.7 PLE downregulates ICP1 capsid expression
(**A**) Percent reads abundance for PLE1, *V. cholerae* chromosomes, and ICP1 over the infection time course. (**B**) Reads normalized to a TPM value based on the total number of reads from each element, and the element's length. (**C**) Volcano plot of ICP1 differential gene expression in the PLE1 culture relative to the PLE(-) culture at 16 minutes post infection. (**D**) Average relative coverage along ICP1's genome in PLE(-) (green) and PLE1 (blue) cultures at 16 minutes post infection. The insert depicts ICP1's head morphogenesis operon. (**E**). Representative Western blot against Gp122, ICP1's major capsid protein from infections of PLE(-), PLE1, PLE2, PLE3, PLE4, and PLE5 cultures at 16 minutes post infection. Quantification and replicates are shown in S17 Fig. For panels A-D, results incorporate values obtained for three biological replicates.

Conclusions

Here, we have provided the first study of phage satellite transcriptomics, obtaining transcriptional programs for ICP1, as well as all five variants of PLE. Aside from broadening our understanding of ICP1 and PLE biology, this work provides surprising insights into the biology of phage satellites. PLE integration alters the gene expression of other mobile elements in *V. cholerae*. Notably, one PLE increases the mobility of CTXϕ, showing that viral satellites can affect the spread of virulence genes that are encoded by unrelated mobile elements. More directly related to the PLE life cycle, we discovered that PLEs do not induce large scale changes to ICP1's transcriptome, suggesting that PLE has adapted to take advantage of ICP1's lytic program as it occurs under conditions permissive to ICP1 replication. The notable exception is that PLEs downregulate ICP1's capsid morphogenesis operon, and this activity is conserved among PLEs, and convergently evolved in the phage satellite P4. This pattern of evolution suggests a strong selective pressure for viral satellites to tune capsid expression, perhaps as a means to optimize resource use for satellite spread. It will be interesting to see if the patterns established here extend to other viral satellites and what other surprising aspects of viral satellite biology will emerge in the future.

Acknowledgements

We thank Andrew Camilli for sharing the CTX minus recipient strain. We thank the University of California Berkeley QB3 Core facility for cDNA library preparation and sequencing. We thank all members of the Seed lab past and present for useful discussions and particularly Kristen LeGault and Caroline Boyd for critical feedback. We also thank Bob Bender, Matt Chapman, and Lyle Simmons for helpful discussions.

Chapter 4

Building a Cohesive Model of PLE parasitism

The results outlined in the previous two chapters shed light on the parasitic life cycles of ICP1 and PLE. As seen in chapter 3, ICP1 dramatically takes over transcription inside the host providing roughly 30% of transcripts by 4 minutes post infection, and upwards of 80% of transcripts by 12 minutes post infection. Late in infection, the majority of ICP1 transcripts are comprised of structural genes, suggesting that a large share of the cell's remaining energy is now going into virion production. This is also reflected by the DNA-seq experiments in chapter 2, that show a sharp increase in ICP1 genome abundance by 12 minutes post infection. By 16 minutes post infection, total ICP1 DNA already exceeds that of the host. These results reveal a parasitic strategy where ICP1 quickly takes over the host cell, and dedicates roughly the second half of infection to virion production. With this strategy, ICP1 is able to produce about 90 new virions in the duration of *V. cholerae*'s doubling time [29]. By contrast, the PLE's takeover of transcription is not nearly as dramatic. PLE genes are robustly expressed, but this does not appear to be at the expense of ICP1's gene expression. Only a single ICP1 gene cluster, the one encoding capsid morphogenesis genes is downregulated, and only about three-fold. PLE has more of an effect on total ICP1 DNA, reducing ICP1 DNA replication three to four-fold while PLE itself is able to replicate over 1000-fold.

It is interesting to consider why ICP1 induces drastic changes on cellular gene expression, while PLE's influence on ICP1 gene expression is minor. If the benefit of parasitism is that it allows the parasite to benefit from the host's own adaptations, then the level of the parasite's manipulation and influence on the host should vary depending on how well the host's life cycle fits the need of the parasite. Infecting *V. cholerae* provides ICP1 with protection from the external environment in the form of the cellular envelope, use of the cell's gene expression machinery, and access to the cell's energy and chemical resources needed for anabolism. While these resources are highly valuable to ICP1, the cell's own goals of division and long term survival are quite distinct from the production of many progeny phage. For this reason, it makes sense that ICP1 would replace cellular gene expression with its own. Dramatic changes are necessary if the cell is to produce many phage in a short amount of time.

By contrast, the PLE's reproductive needs overlap nicely with ICP1's. For the PLE to effectively reproduce from an infected cell, there needs to be replication of many PLE genomes, production of virion structural components, and lysis of the cell. ICP1's life cycle is mostly already suitable for this aim. PLE simply needs to redirect replication and virion packaging to its own genome to propagate. Recent work by my colleagues shows that the PLE does remodel virion capsids to better fit its genome (T. Silvas, unpublished data). This activity is consistent with the PLE's reduction of ICP1 capsid expression. By reducing the amount of capsid made relative to phage tails, PLE is altering ICP1's reproductive program to better fit the PLE's needs.

These adaptations for PLE horizontal mobility are at odds with PLE's poor transduction efficiency that has been previously reported [29], and is cited in chapter 2. Recently, it has been found that PLE transduction efficiency increases by more than two orders of magnitude if growth media is supplemented with magnesium (K. LeGault, unpublished data). This result makes sense in light of PLE's capsid remodeling activities. Different

ion requirements for capsid stability are a likely explanation for why PLE transduction appears inefficient in the absence of magnesium. Capsid remodeling, along with the more robust transduction under certain media conditions, suggests that the horizontal mobility remains selected for in the PLE, and is an important part of the PLE's life cycle.

Two of the biggest questions raised by the results in chapter 2, why does the midi PLE not replicate to the same level as PLE, and what is the nature of the inverse relationship of PLE and ICP1 replication, have been at least partially answered through recent work. PLE has been found to encode an endonuclease that cuts ICP1's genome in a site-specific manner. This endonuclease is partially inhibitory towards ICP1 reproduction and genome replication, and boosts the replication of the midi PLE (K. LeGault, unpublished data). These results point to a model where the PLE is inhibitory towards ICP1 replication through the action of an endonuclease, and the resulting reduction in ICP1 replication frees up resources for the PLEs own replication. The lack of DNA replication interference against ICP1 when PLE's replication initiator is knocked out could be because not enough endonuclease can be produced from a single copy of PLE's genome, or that the PLE will not express any endonuclease without first passing a replication checkpoint. It is interesting to consider if endonuclease cleavage of ICP1 copies generated by rolling circle replication could fragment the new genomes and lead to a loss of linearization as assessed by deep sequencing. While such a scenario seems plausible, the fact that the Δ*repA* PLE losses interference of the ICP1 DNA replication level but retains the block of ICP1 linearization suggests that these two effects stem from separate PLE processes, though it is possible that different levels of PLE endonuclease production could have different effects on ICP1's genome.

The loss of PLE genome linearization along with the remodeling of capsids suggest that the blocking of genome packaging is the primary mechanism of ICP1 restriction (Fig 3.1). This is consistent with what has been found in other bacteriophage satellite systems, where capsids are remodeled, and, in the case of some SaPIs, genome packaging machinery is repressed and replaced with proteins with specificity for the phage satellite's genome [44,137]. For the PLE, auxiliary mechanisms that reduce ICP1 DNA replication and capsid production likely contribute to ICP1 restriction while boosting PLE reproduction. In fact, all of PLE's inhibitory activities against ICP1 that have been discovered so far can be inferred to confer some benefit to PLE mobilization. The PLE appears to have evolved to restrict ICP1 infection among the cells it already inhabits, while also spreading to new host strains.

An important caveat to remember when thinking about PLE and ICP1's relationship is that within the existing library of isolates, ICP1 isolates typically have adaptations, such as CRISPR-Cas, that allows them to overcome contemporaneous circulating PLEs [29]. This may be partially due to limited sampling, and there may be cyclical changes in PLE's and ICP1's relative level of success that are driven by new adaptations. It is also important to consider that CRISPR restriction of PLE does allow some level of PLE particle production, even though ICP1 is not restricted [29]. The phage satellite P4, is much more successful when infecting lysogens of the phage it parasitizes than it is when responding to infection by that phage, because P4 can activate that phage's late

gene expression and bypass the phage's replication [39,138]. In a manner similar to P4, PLE could possibly be capable of secondary infections of cells already infected by ICP1. While ICP1 is not known to lysogenize its host, it has been found to undergo prolonged reproduction in a lysis inhibited state [97]. Since PLE appears to mostly rely on late ICP1 infection processes for its life cycle, perhaps an activated PLE could secondarily infect a lysis inhibited ICP1 cell and produce additional PLE particles. This could explain why PLE has been found to inhibit lysis inhibition by invading ICP1 [97]. By accelerating cell lysis and hastening its release, PLE might be able to spread among lysis inhibited cells when the concentration of phage is relatively high. While the work in this thesis provides good insight into the mechanisms of PLE parasitism, the results show a narrow snapshot of the myriad of possible infection conditions. In addition to finding and defining more PLE activities, PLE and ICP1 interactions will have to be explored in different contexts to better understand their parasitic life cycles.

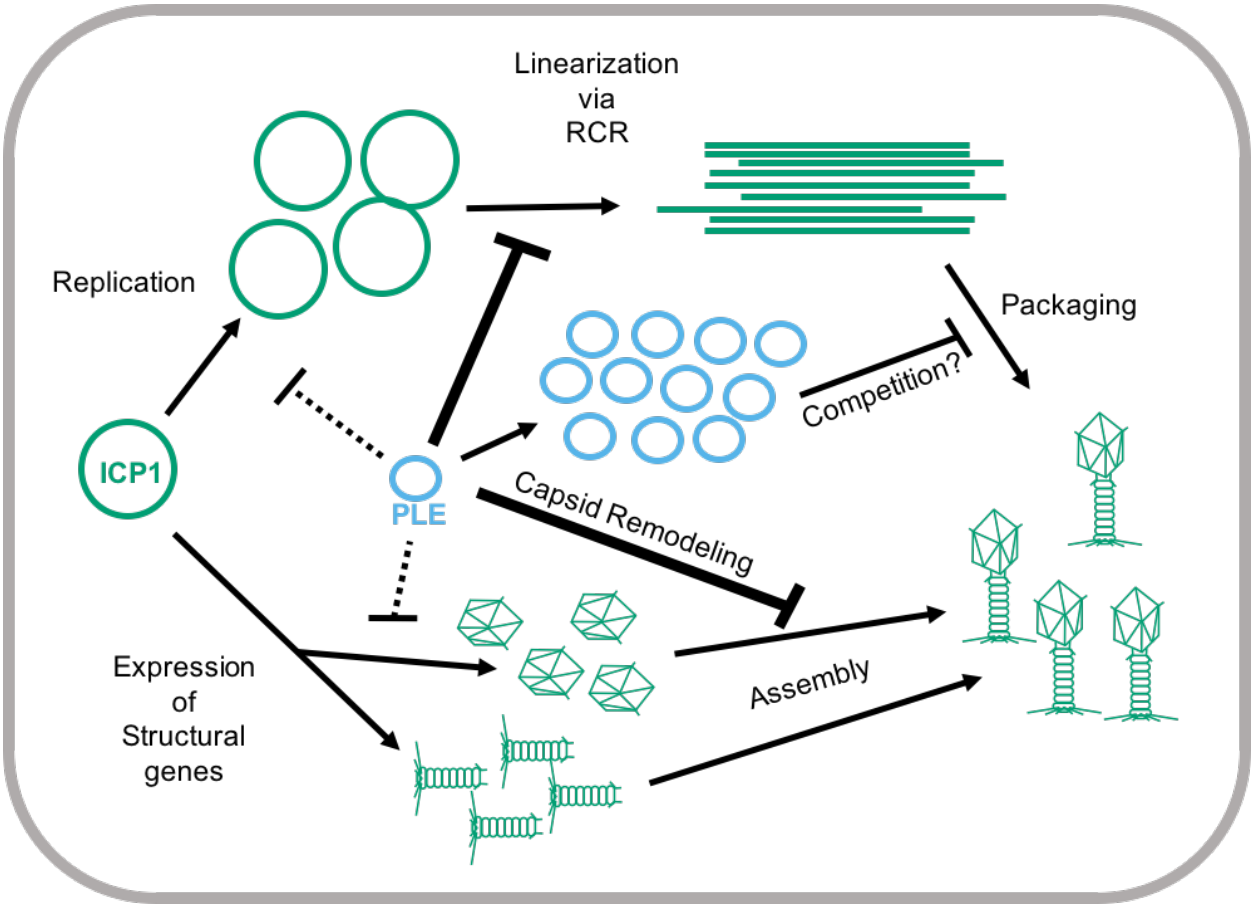
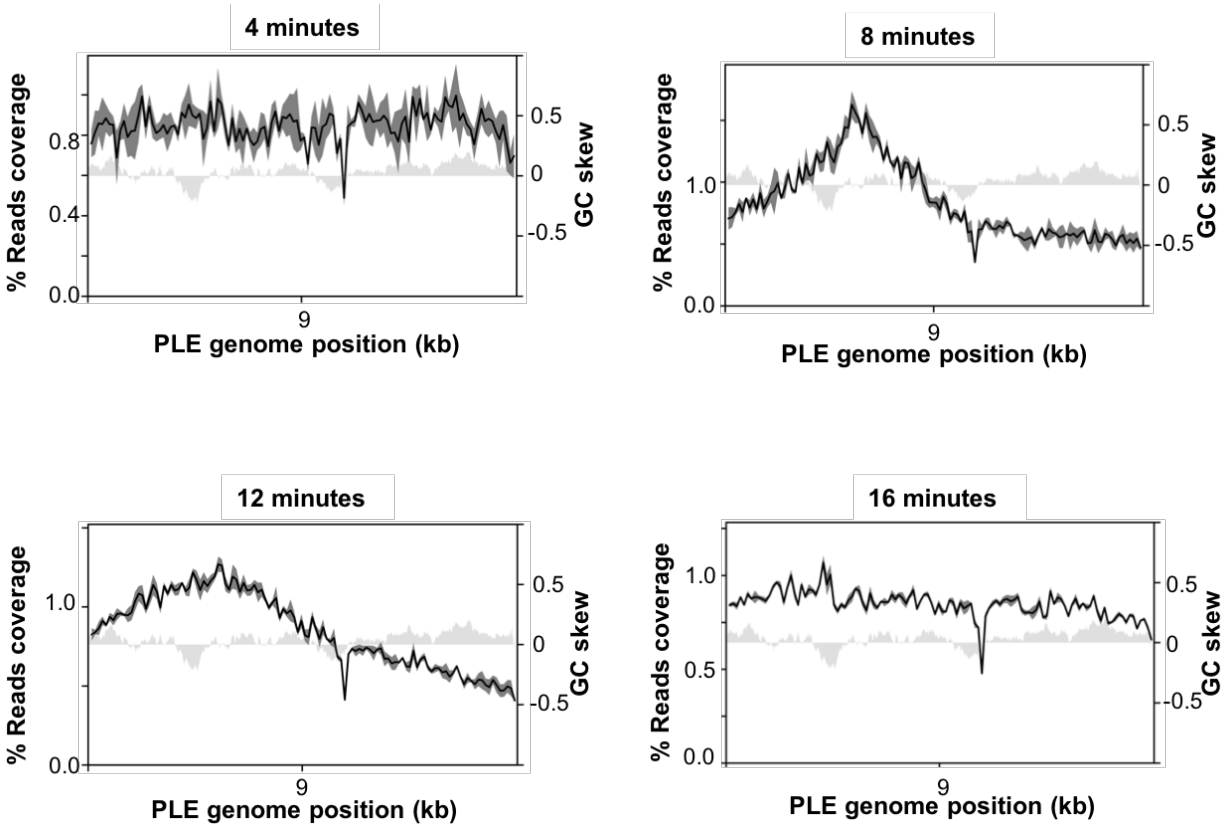
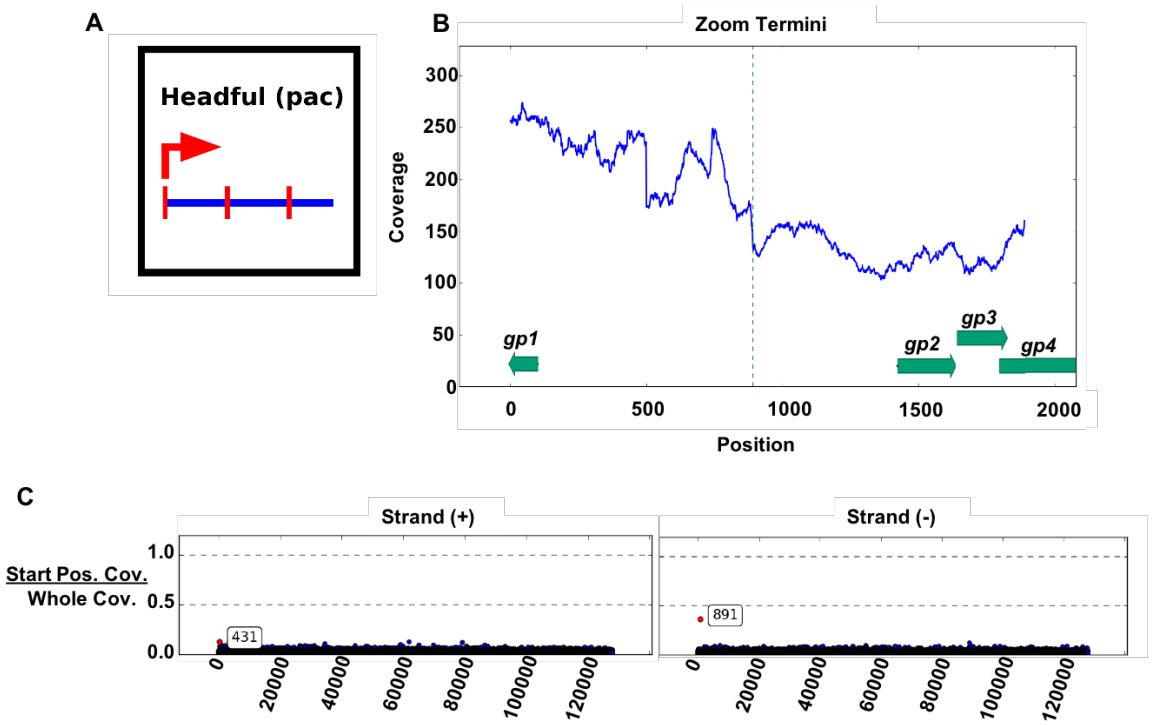


Figure 3.1 Mechanisms of ICP1 restriction by PLE. A diagram of ICP1's life cycle showing which steps are interfered with by the PLE. Dotted lines indicate that the observed interference activity is not predicted to be wholly restrictive to ICP1, while bolded lines indicate that the observed interference activity is expected to broadly prevent ICP1 virion production.

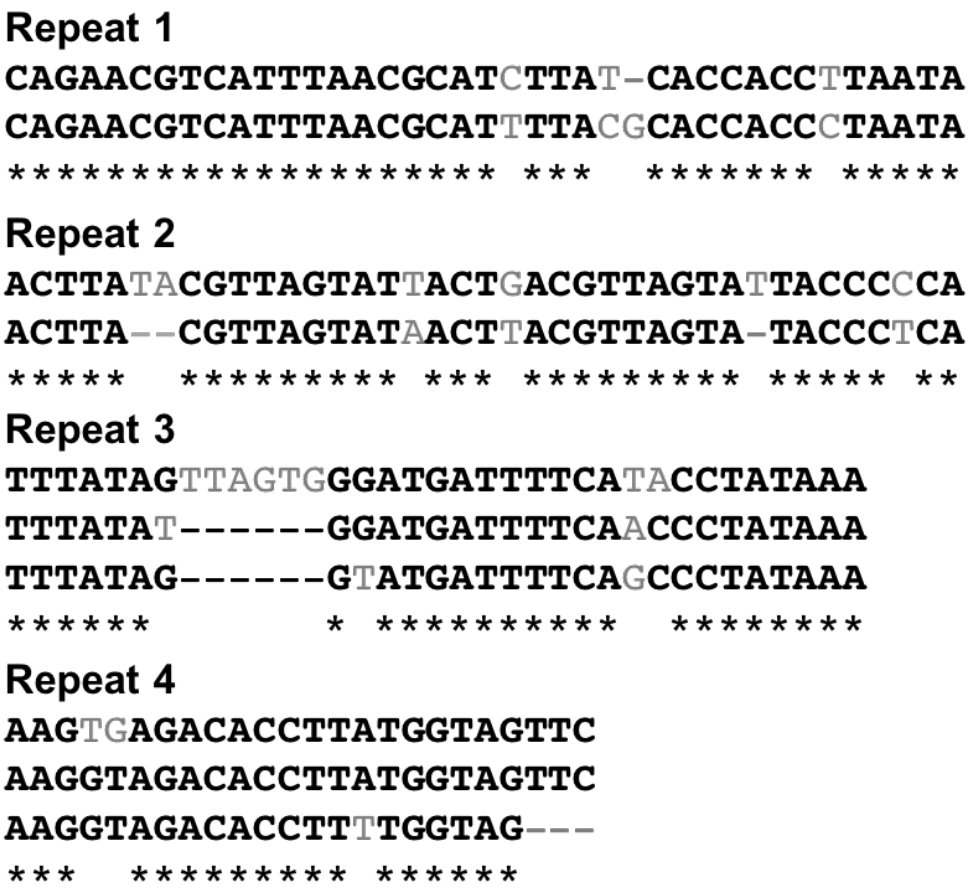
Supplementary Information



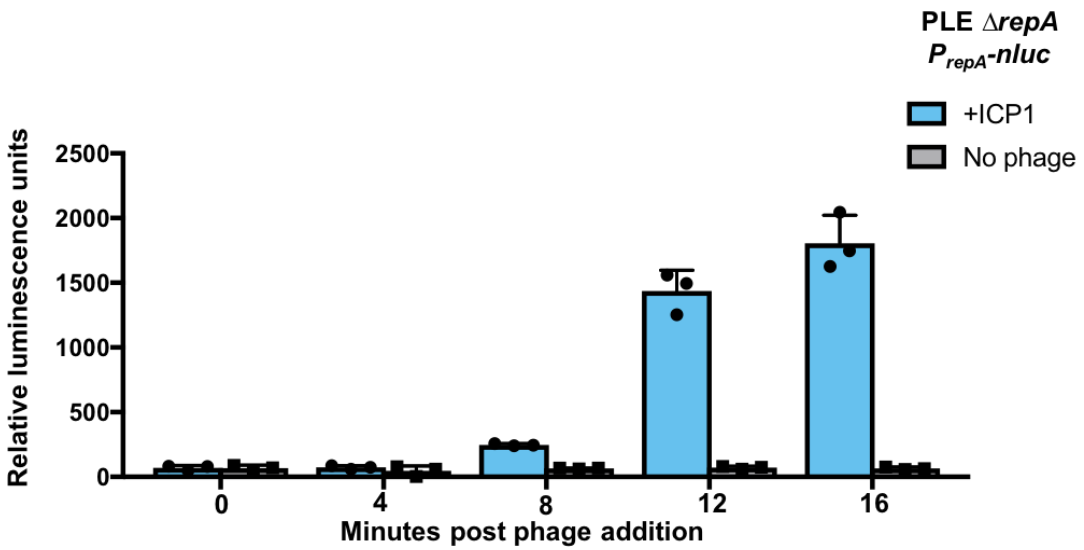
Supplementary Figure 1.1. Percent reads coverage plots across the PLE genome during ICP1 infection. For each time point, the percent reads coverage across the genome for three biological replicates was determined. The average percent reads coverage is shown as a black line, while standard deviation appears as dark gray shading around the line. The GC skew is plotted on the right axis as light gray shading.



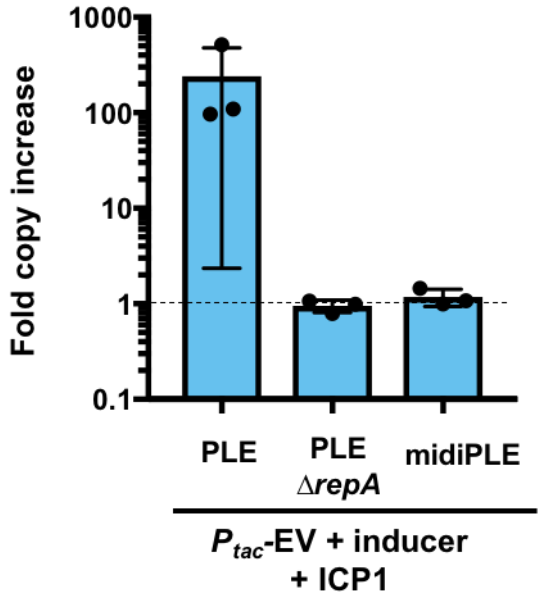
Supplementary Figure 1.2. ICP1 is predicted to use a headful packaging mechanism dependent on a pac site as determined by PhageTerm analysis. **(A)** PhageTerm schematic showing the predicted packaging mode of ICP1. ICP1 DNA is packaged into capsids using a headful mechanism from a distinct site on the phage genome. **(B)** A zoomed in view of ICP1's packaging terminus with whole genome coverage plotted. **(C)** Plots of reads start position coverage divided by whole coverage along the entire ICP1 genome. The (+) strand (left) and (-) strand (right) are plotted separately.



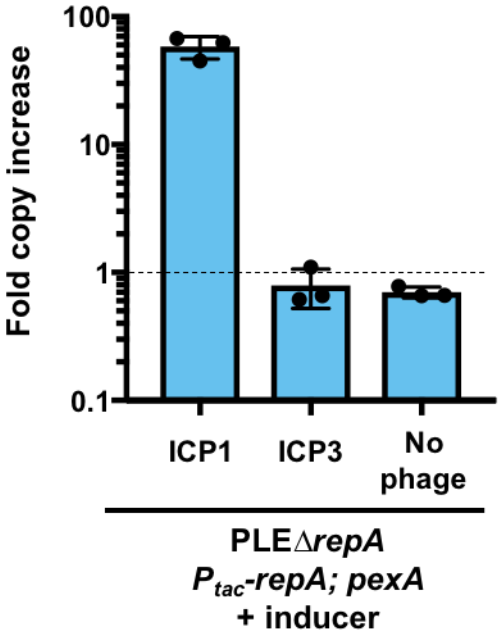
Supplementary Figure 1.3. Repeat sequences found within the PLE noncoding region. Mismatches are shown in light gray, asterisks represent conserved sequence. Repeats 1 and 2 are interspersed with each other across a 528bp region (Figure 3B). The repeat 3 sequences are proximal to each other separated by 5 and 14bp, and the repeat 4 sequences are contiguous.



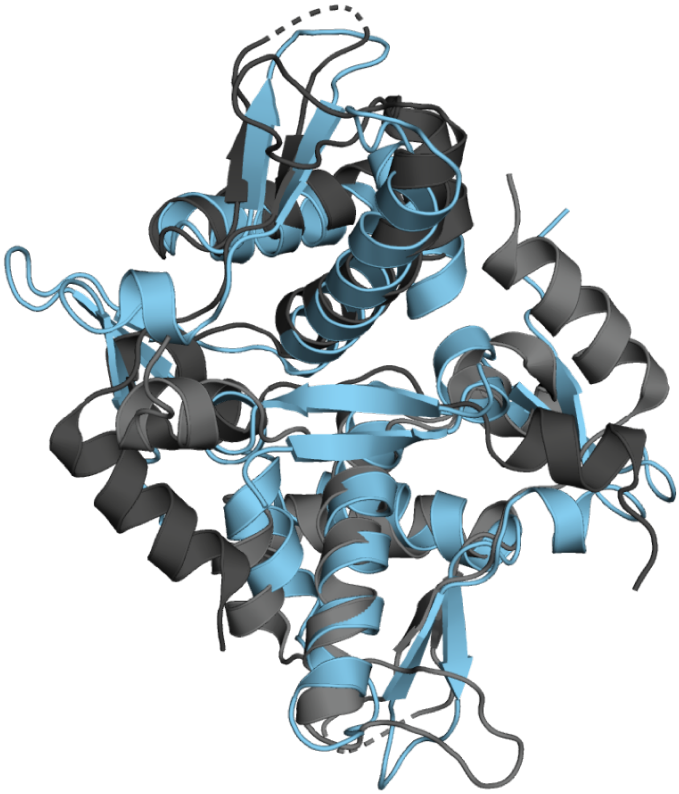
Supplementary Figure 1.4. Relative luminescence units of a PLE PrepAnluc reporter strain, where the repA gene has been replaced by nanoluciferease (nluc). Values shown are the means of three biological replicates.



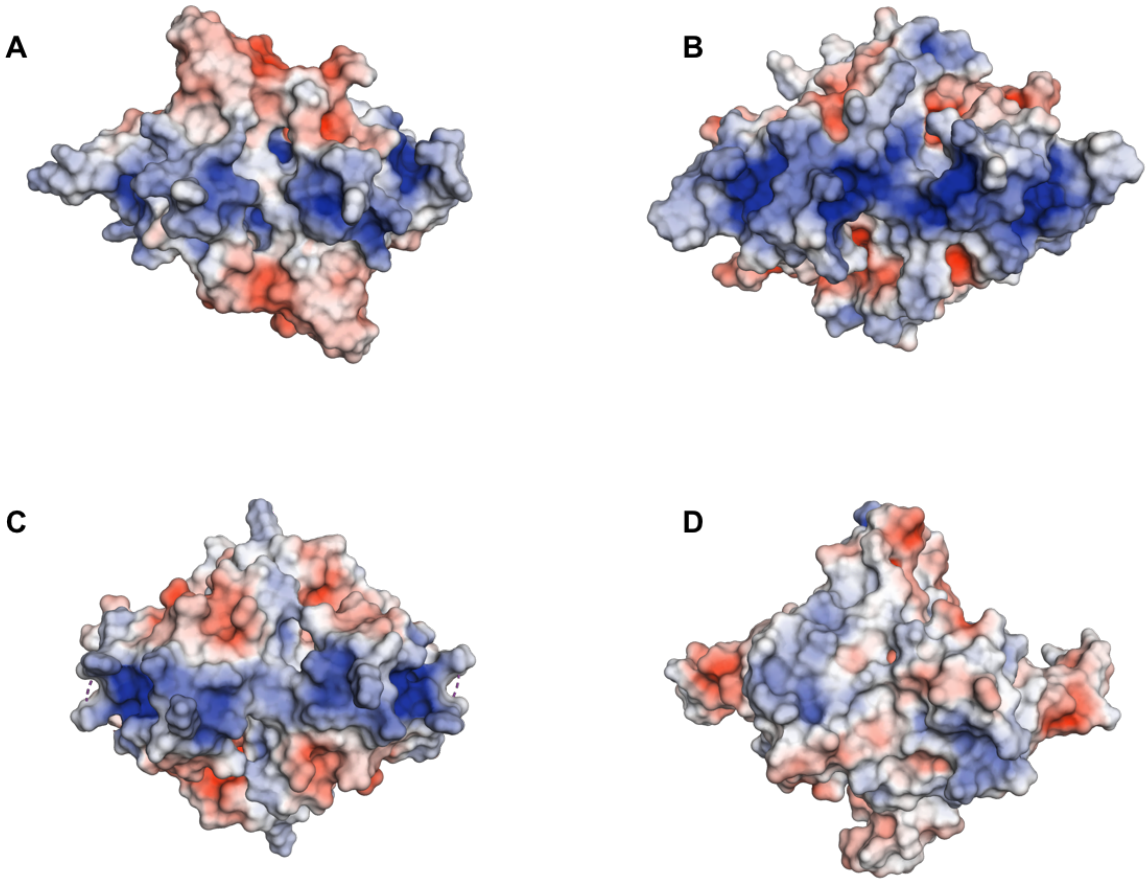
Supplementary Figure 1.5. Fold copy increase of wild-type PLE, PLE Δ*repA*, and the midiPLE following induction of an empty vector control (EV) 20 minutes post-infection with ICP1 as assessed by qPCR. The dashed line indicates no change in PLE copy



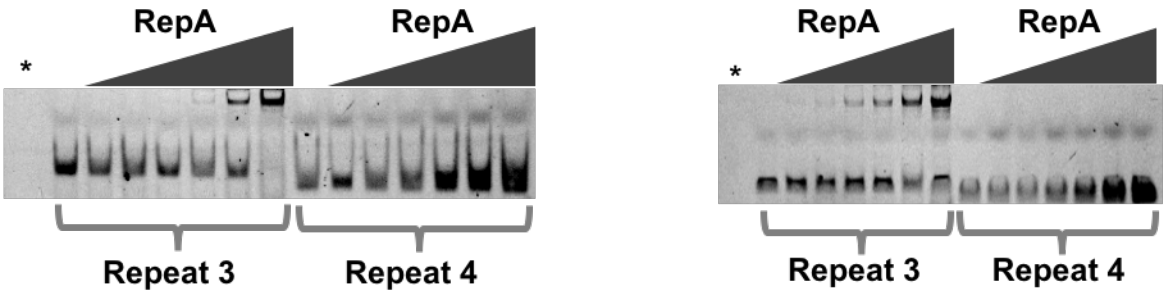
Supplementary Figure 1.6. Fold copy increase of PLE Δ*repA*, following induction of RepA and PexA expression and addition of ICP1, ICP3, or buffer control 20 minutes post-addition as assessed by qPCR. The dashed line indicates no change in PLE copy



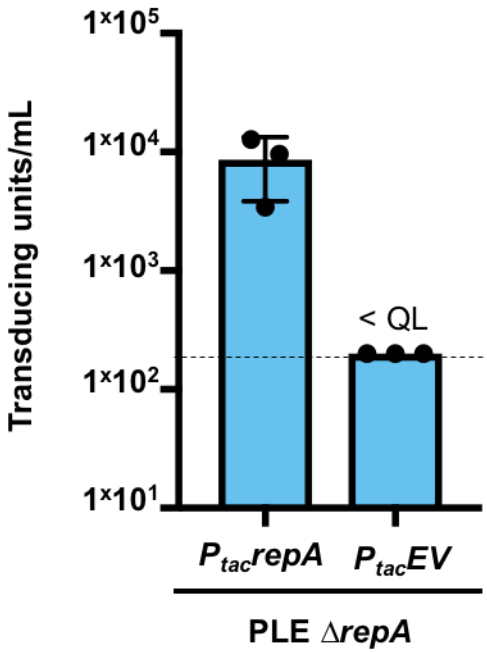
Supplementary Figure 1.7. Overlay of ribbon diagrams for the crystal structures of the NTD of dimers of PLE RepA (light blue) and pSK41 (dark grey) (RMSD = 4.197942 over 184 residues).



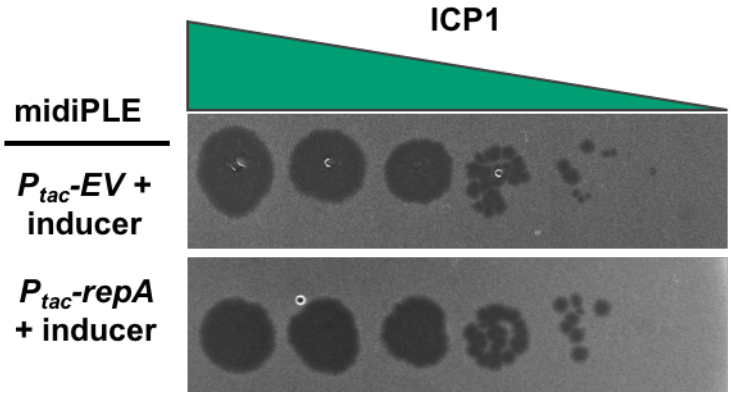
Supplementary Figure 1.8. Electrostatic profiles for PLE RepA (A), pTZ2162 RepA (B), pSK41 (C), and B. subtilis DnaD (D), N-terminal domain dimers. Positive (blue) and negative (red) charges are indicated on the surface.



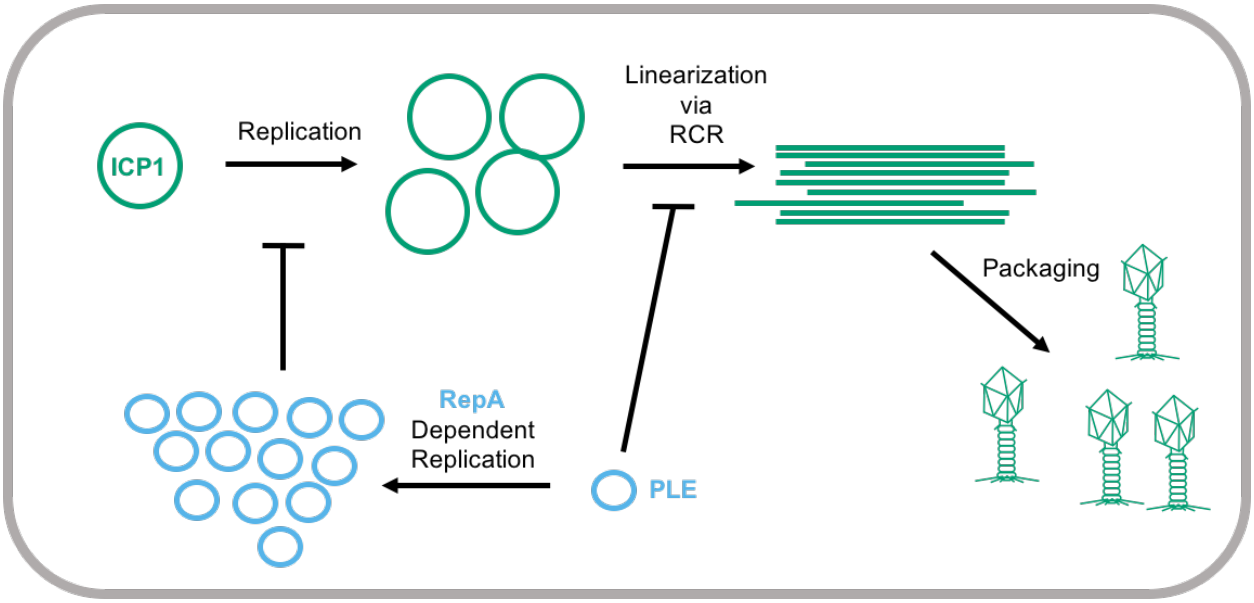
Supplementary Figure 1.9. Replicates of an electrophoretic mobility shift assay using probes from the PLE noncoding region. RepA binding was tested for probes corresponding to the repeat 3 or repeat 4 sequence from the PLE NCR3. The * denotes a RepA(+) DNA (-) control



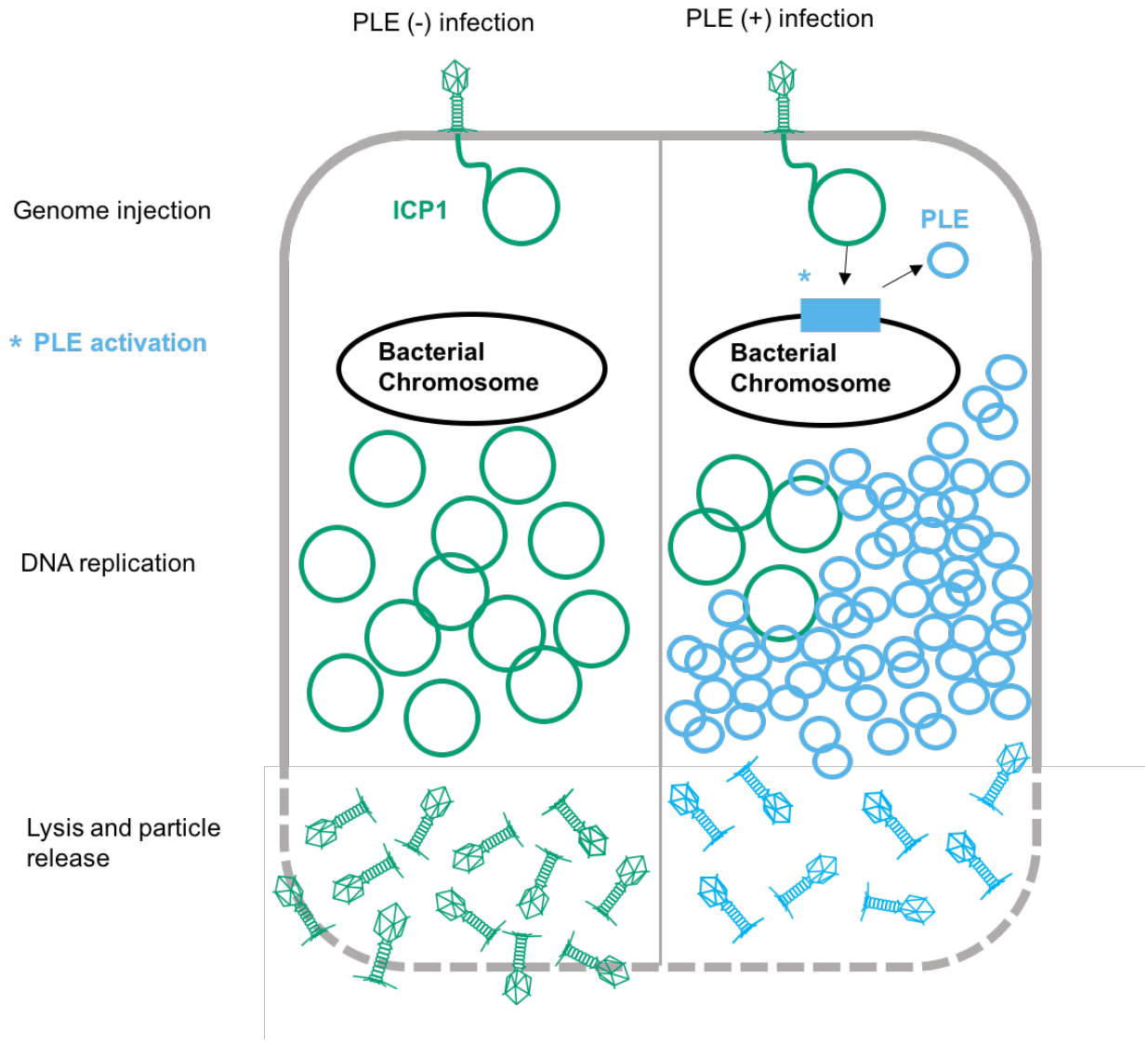
Supplementary Figure 1.10. PLE replication aids transduction. Transduction units per mL produced from ICP1 infection of Δ*repA* PLE complemented with repA or an empty vector control (EV). Quantification limit (QL) = 200 TU/mL.



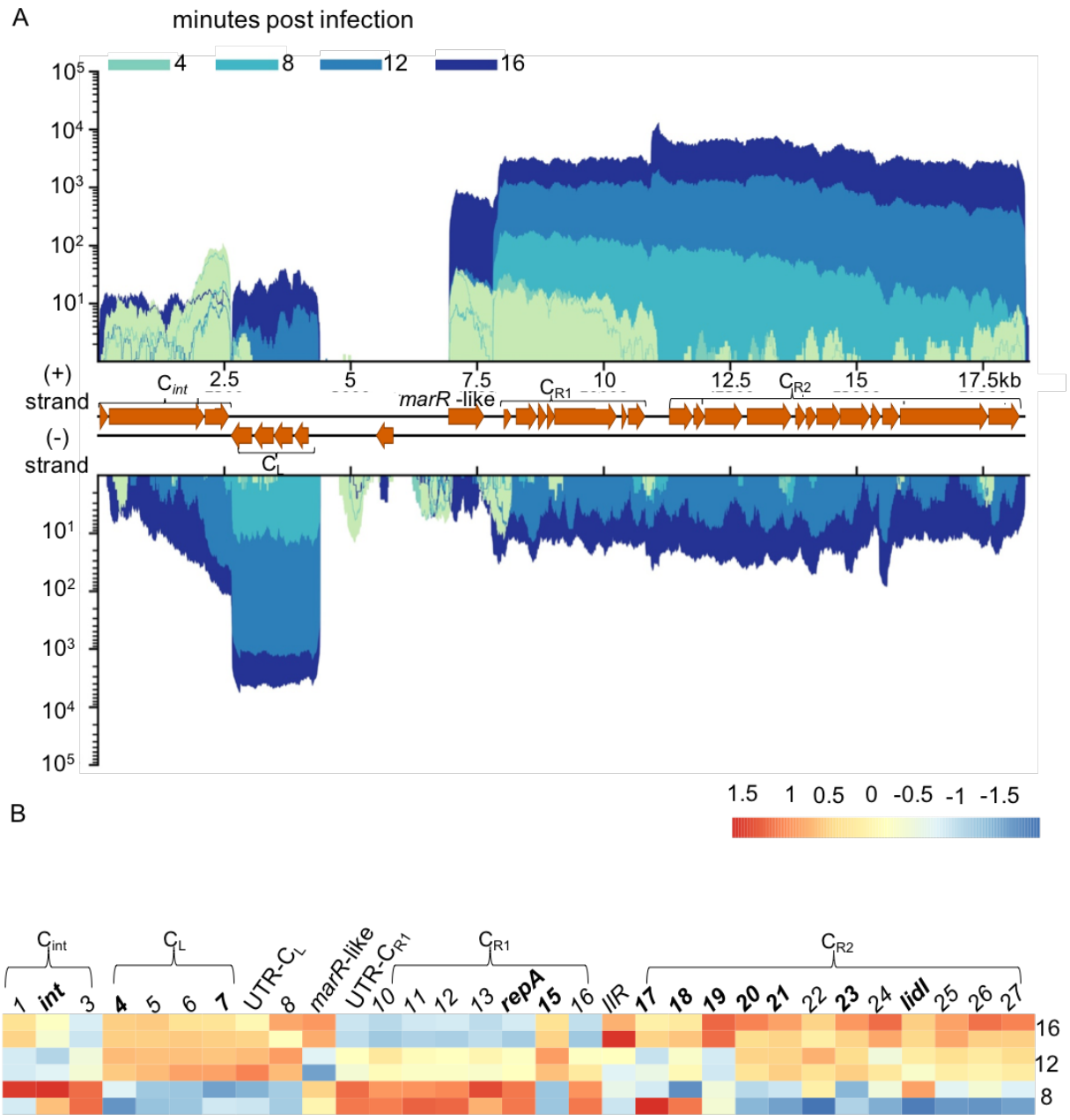
Supplementary Figure 1.11. Spot assay showing ICP1 susceptibility for *V. cholerae* harboring the midiPLE complemented with an empty vector (EV) or repA. Serial tenfold ICP1 dilutions spotted onto PLE (+) and PLE (-) *V. cholerae* lawns (grey). Zones of cell death are shown in black.



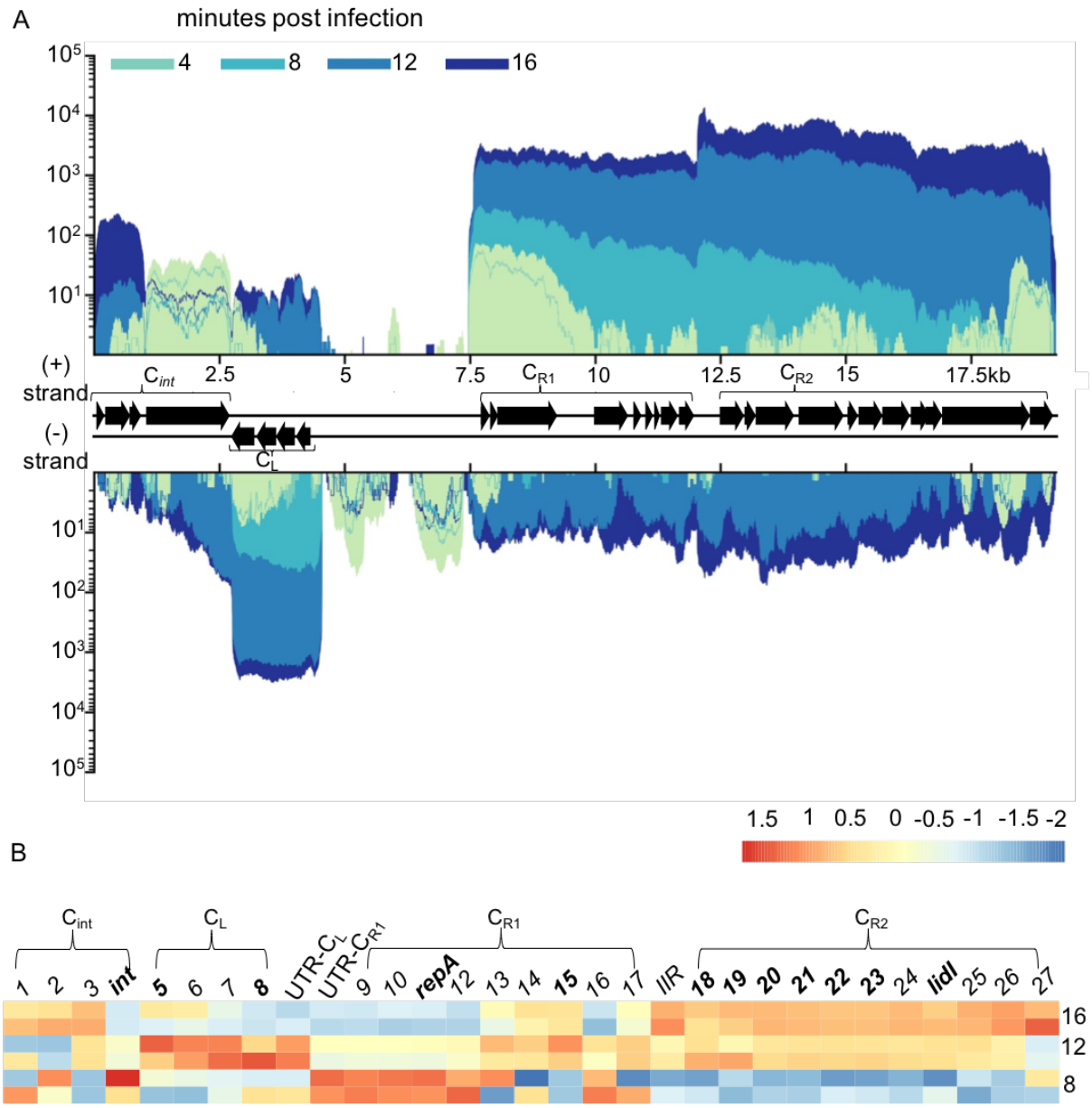
Supplementary Figure 1.12. Model of PLE interference of ICP1 replication. ICP1 begins replication through a bidirectional theta mechanism before switching to rolling circle replication (RCR). RCR linearizes the ICP1 genome so that it can be packaged into capsids. The PLE is able to block ICP1 linearization without replicating. PLE inhibition of ICP1 copy increase is dependent on PLE replication.



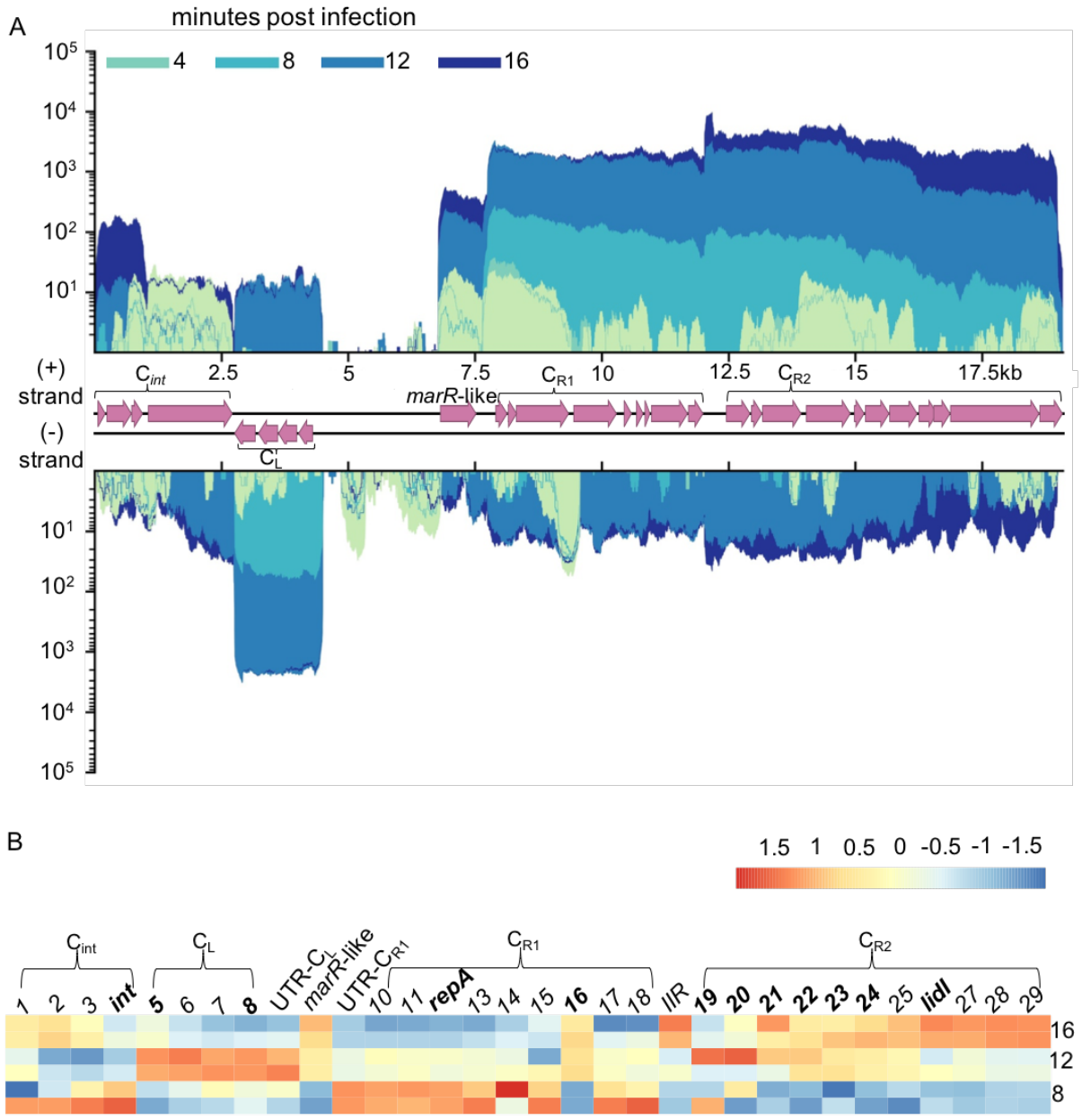
Supplemental Figure 2.1 PLE life cycle Model of ICP1 infection in PLE(-) and PLE(+) *V. cholerae*. ICP1 injects its DNA into *V. cholerae*; prior to DNA replication, ICP1 activity leads to PLE activation and excision. ICP1 DNA replication is reduced in the PLE(+) cell where the PLE replicates to high copy. Finally, the cell lyses and releases infectious particles. No ICP1 particles and of PLE transducing particles are released from the PLE(+) cell.



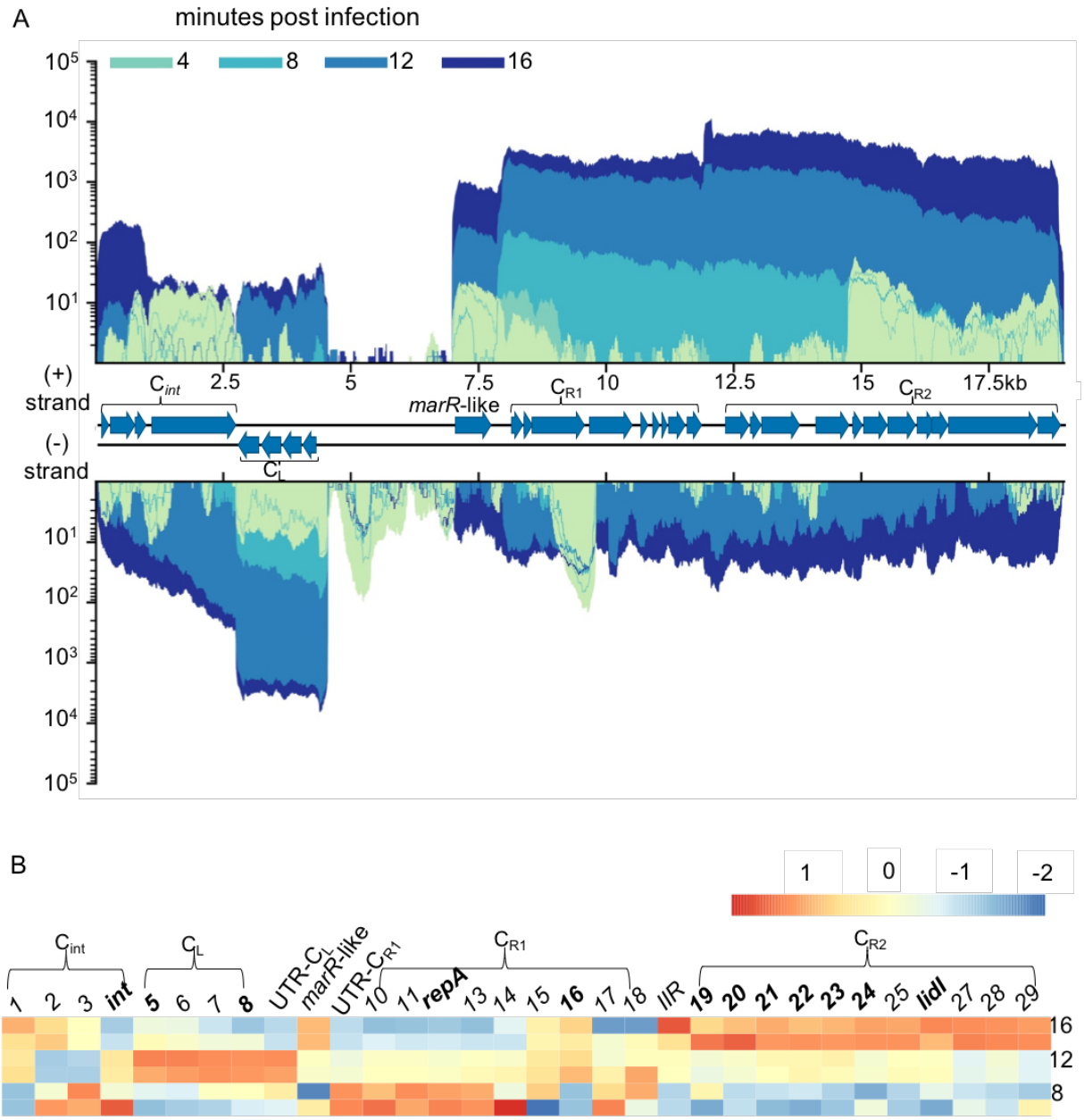
Supplemental Figure 2.2 PLE2 transcriptome (**A)** PLE2's genome displaying average reads coverage over the course of infection. Reads are depicted on a logarithmic scale, to improve visibility of early expressed genes. Reads are color coded by time point. (**B**) Heat map of PLE1 gene expression over the course of infection. Color reflects the Z-score of each gene's log2 TPM value across replicates and time points. *V. cholerae* and ICP1 genes were excluded from TPM calculation to highlight relative changes in PLE1 transcript abundance. Core genes, protein coding genes with high conservation across PLEs, are bolded. Values for 8, 12, and 16 minutes post infection are shown. Results incorporate gene expression values obtained for two biological replicates.



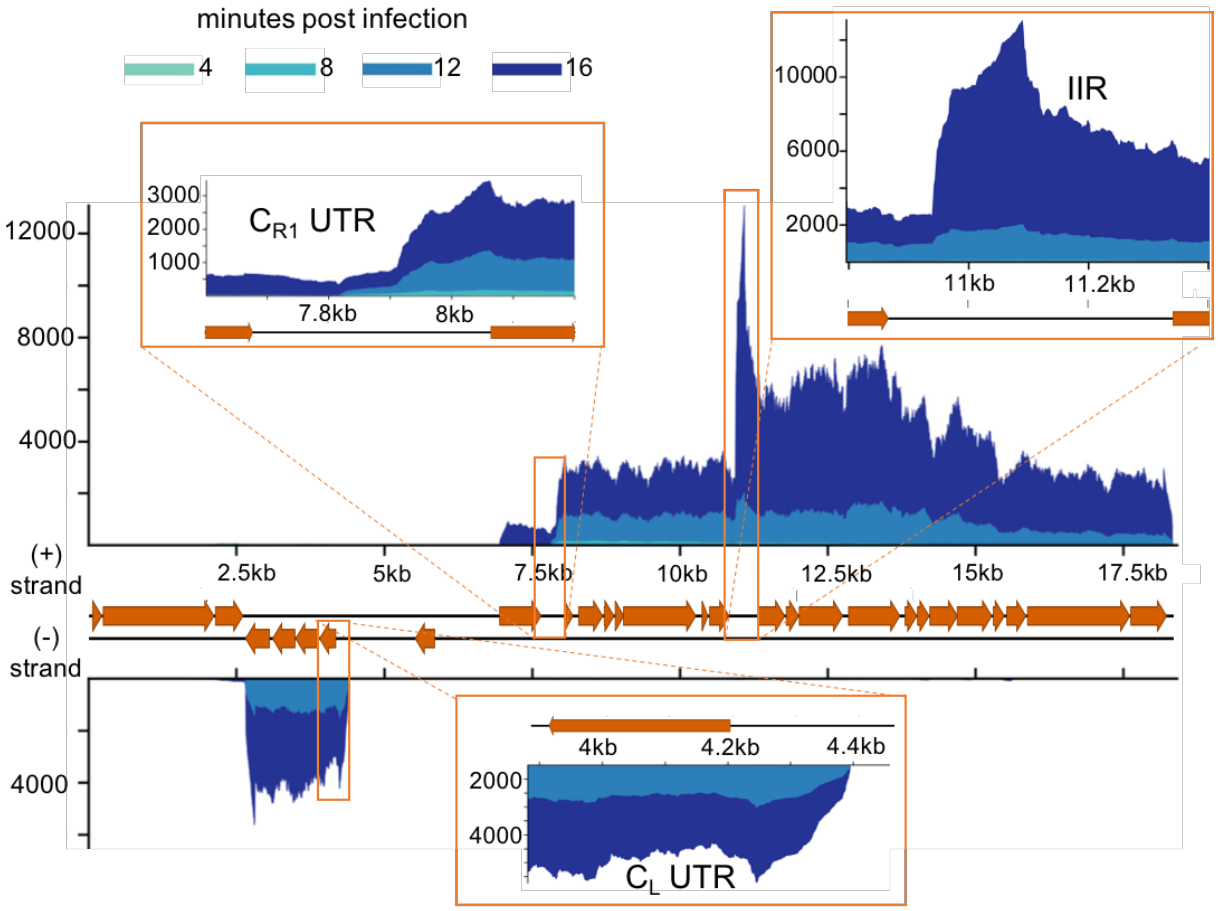
Supplemental Figure 2.3 PLE3 Transcriptome (**A)** PLE3's genome displaying average reads coverage over the course of infection. Reads are depicted on a logarithmic scale, to improve visibility of early expressed genes. Reads are color coded by time point. (**B**) Heat map of PLE1 gene expression over the course of infection. Color reflects the Z-score of each gene's log2 TPM value across replicates and time points. *V. cholerae* and ICP1 genes were excluded from TPM calculation to highlight relative changes in PLE1 transcript abundance. Core genes, protein coding genes with high conservation across PLEs, are bolded. Values for 8, 12, and 16 minutes post infection are shown. Results incorporate gene expression values obtained for two biological replicates.



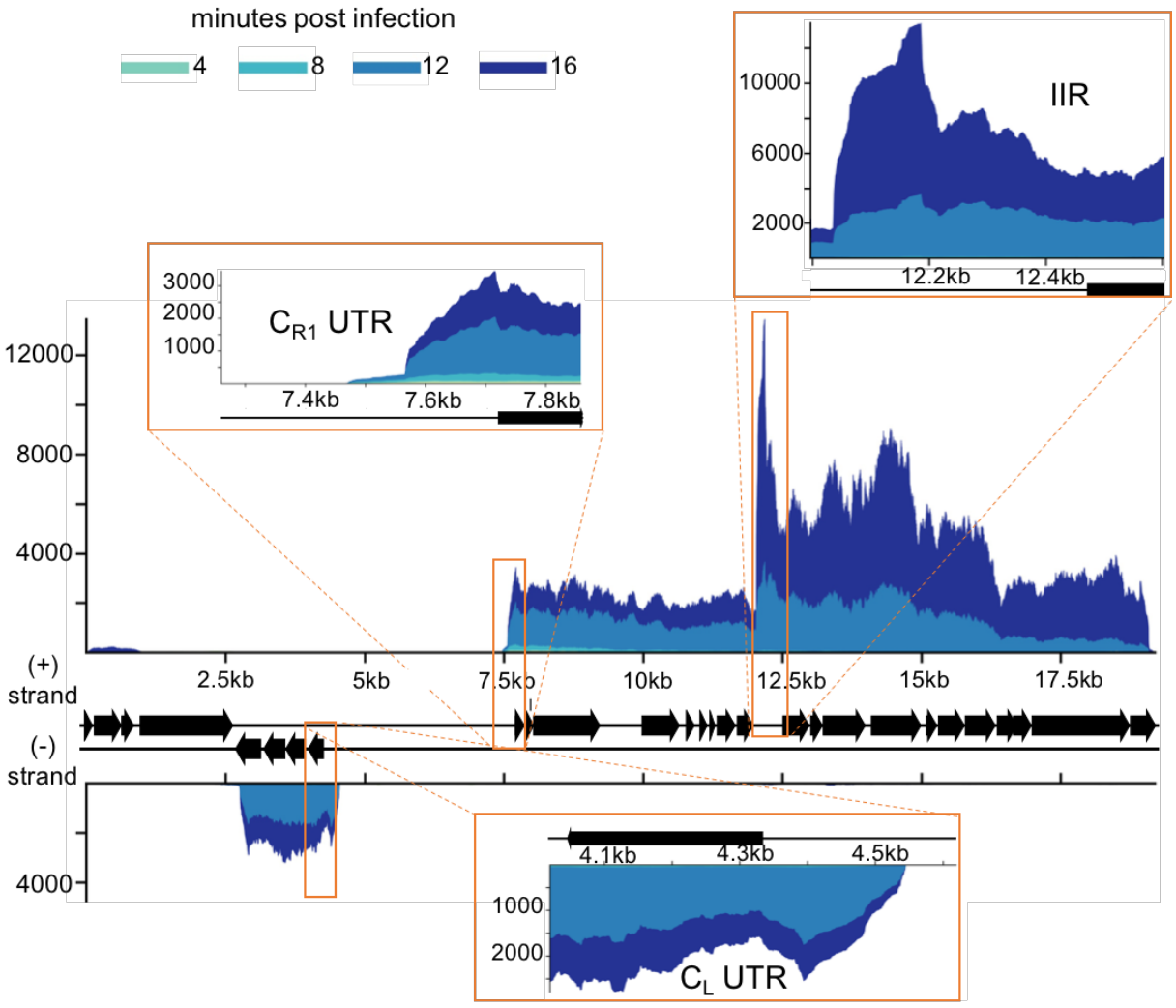
Supplemental Figure 2.4 PLE4 Transcriptome (**A)** PLE4's genome displaying average reads coverage over the course of infection. Reads are depicted on a logarithmic scale, to improve visibility of early expressed genes. Reads are color coded by time point. (**B**) Heat map of PLE1 gene expression over the course of infection. Color reflects the Z-score of each gene's log2 TPM value across replicates and time points. *V. cholerae* and ICP1 genes were excluded from TPM calculation to highlight relative changes in PLE1 transcript abundance. Core genes, protein coding genes with high conservation across PLEs, are bolded. Values for 8, 12, and 16 minutes post infection are shown. Results incorporate gene expression values obtained for two biological replicates.



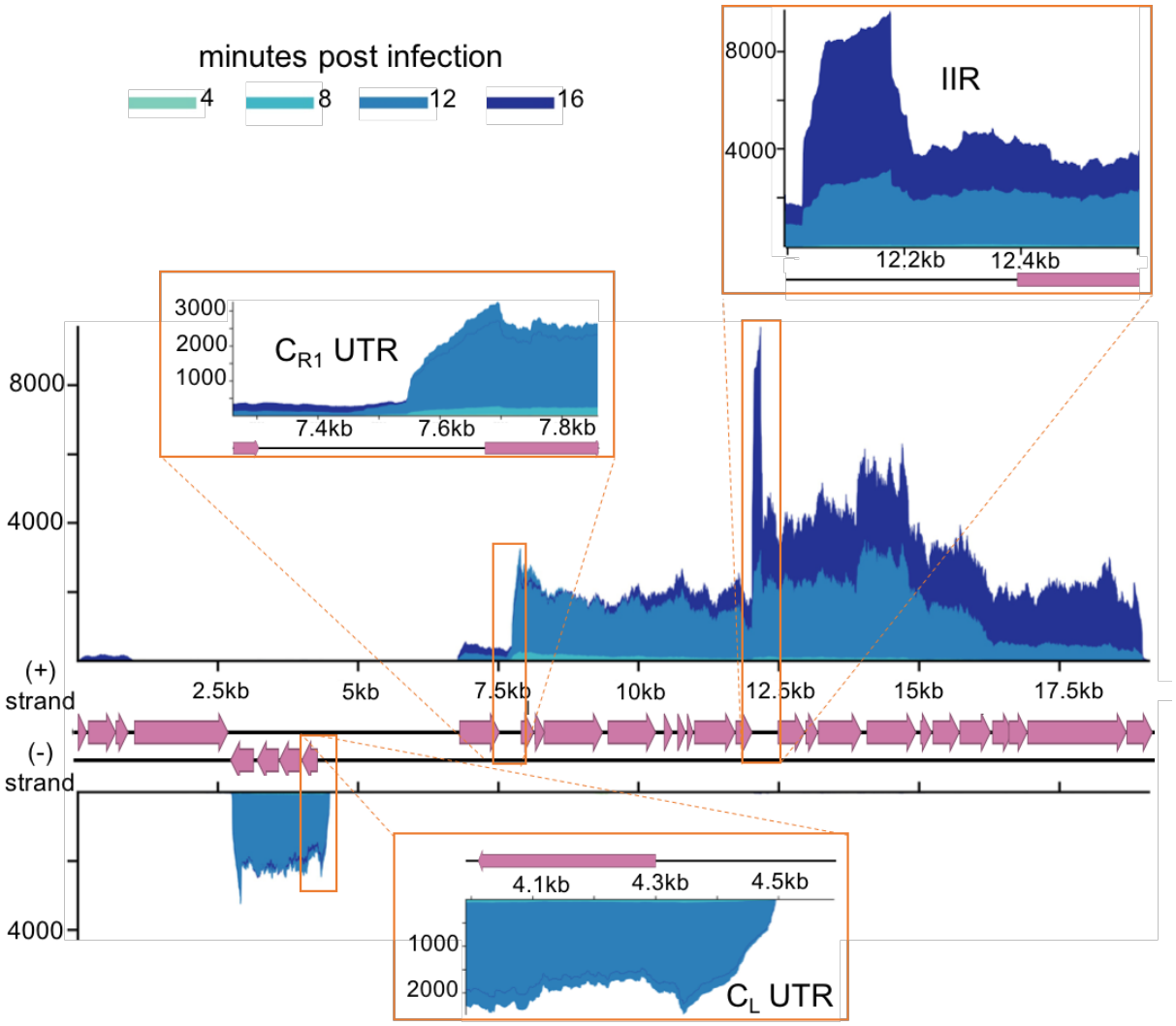
Supplemental Figure 2.5 PLE5 Transcriptome (A) PLE5's genome displaying average reads coverage over the course of infection. Reads are depicted on a logarithmic scale, to improve visibility of early expressed genes. Reads are color coded by time point. **(B)** Heat map of PLE1 gene expression over the course of infection. Color reflects the Z-score of each gene's log2 TPM value across replicates and time points. *V. cholerae* and ICP1 genes were excluded from TPM calculation to highlight relative changes in PLE1 transcript abundance. Core genes, protein coding genes with high conservation across PLEs, are bolded. Values for 8, 12, and 16 minutes post infection are shown. Results incorporate gene expression values obtained for two



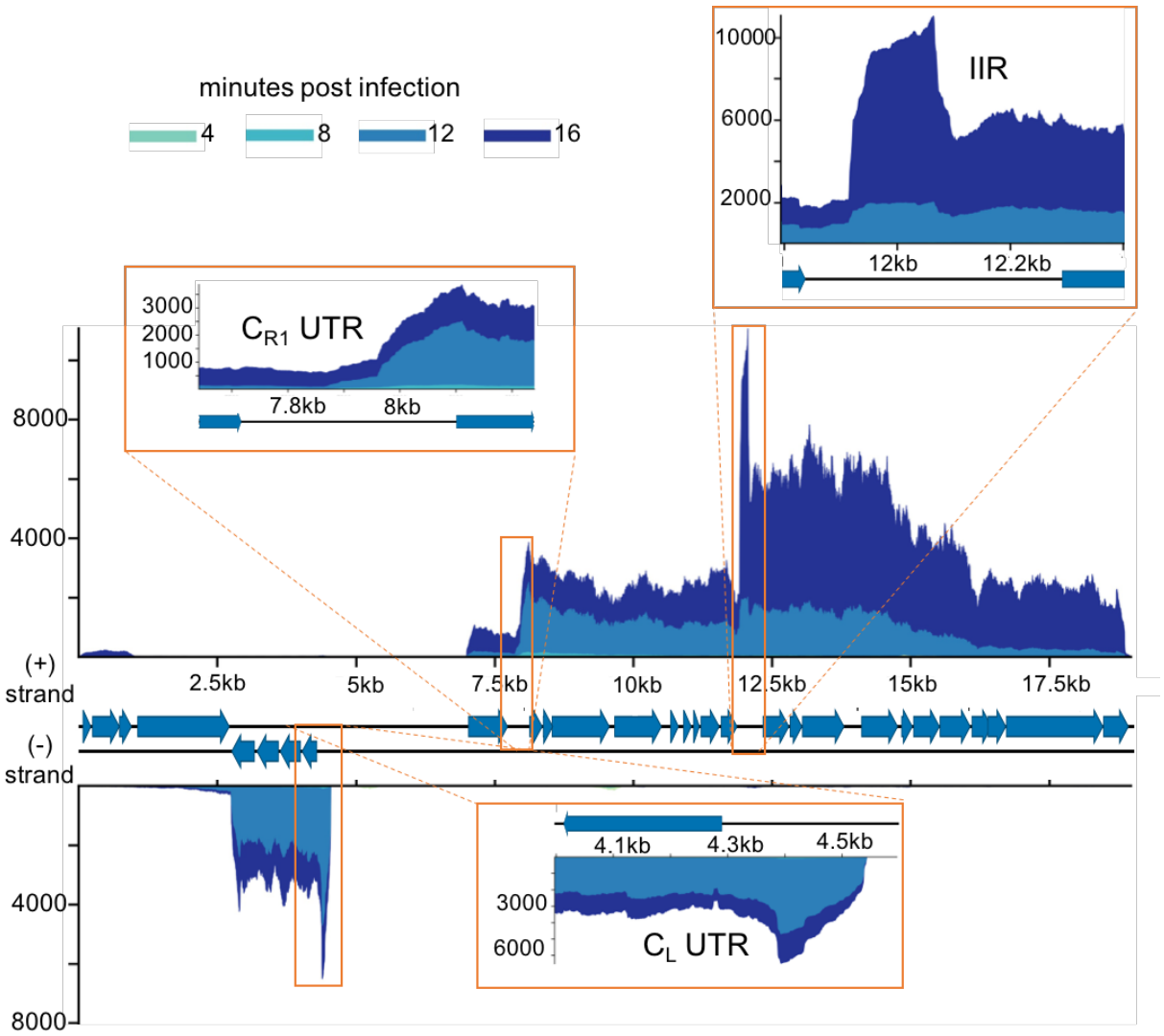
Supplemental Figure 2.6 PLE2 noncoding RNA PLE2's genome displaying average reads coverage over the course of infection on a linear scale. Reads are color coded by time point. Inserts depict detected transcripts that lack predicted coding sequence. Results incorporate gene expression values obtained for two biological replicates.



Supplemental Figure 2.7 PLE3 noncoding RNA PLE3's genome displaying average reads coverage over the course of infection on a linear scale. Reads are color coded by time point. Inserts depict detected transcripts that lack predicted coding sequence. Results incorporate gene expression values obtained for two biological replicates.



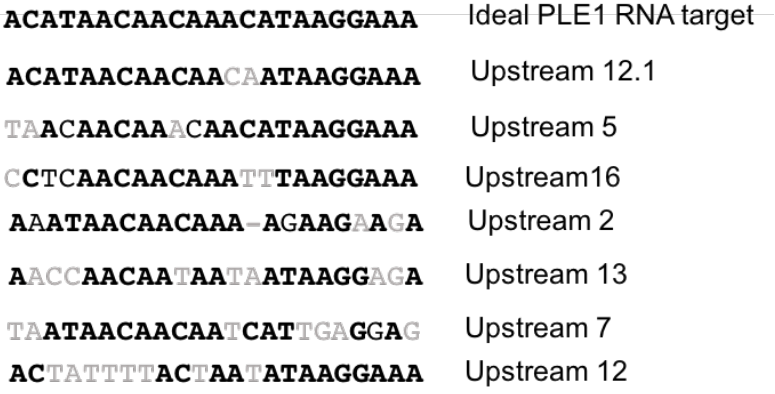
Supplemental Figure 2.8 PLE4 noncoding RNA PLE4's genome displaying average reads coverage over the course of infection on a linear scale. Reads are color coded by time point. Inserts depict detected transcripts that lack predicted coding sequence. Results incorporate gene expression values obtained for two biological replicates.



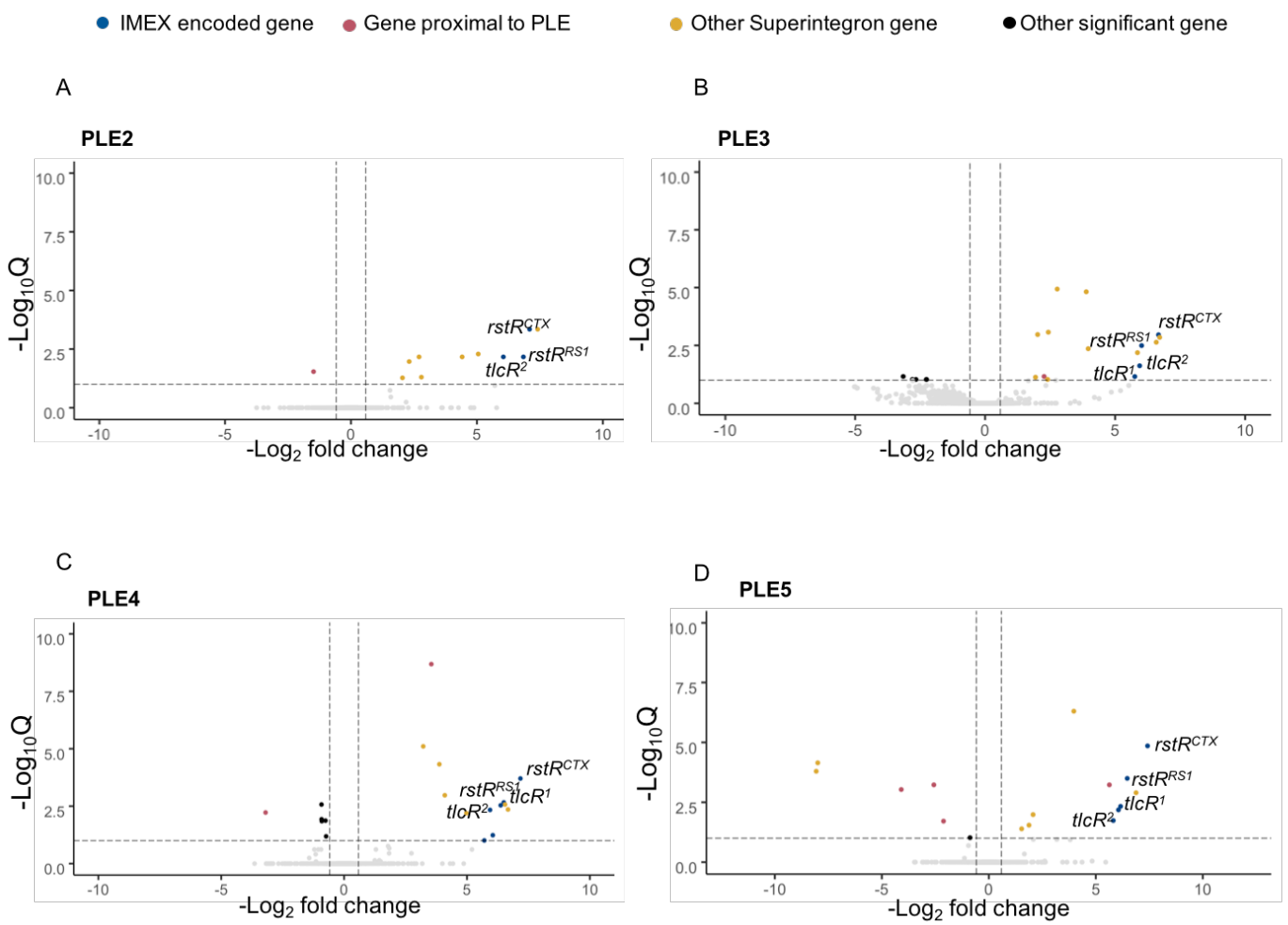
Supplemental Figure 2.9 PLE5 noncoding RNA PLE5's genome displaying average reads coverage over the course of infection on a linear scale. Reads are color coded by time point. Inserts depict detected transcripts that lack predicted coding sequence. Results incorporate gene expression values obtained for two biological replicates.

A

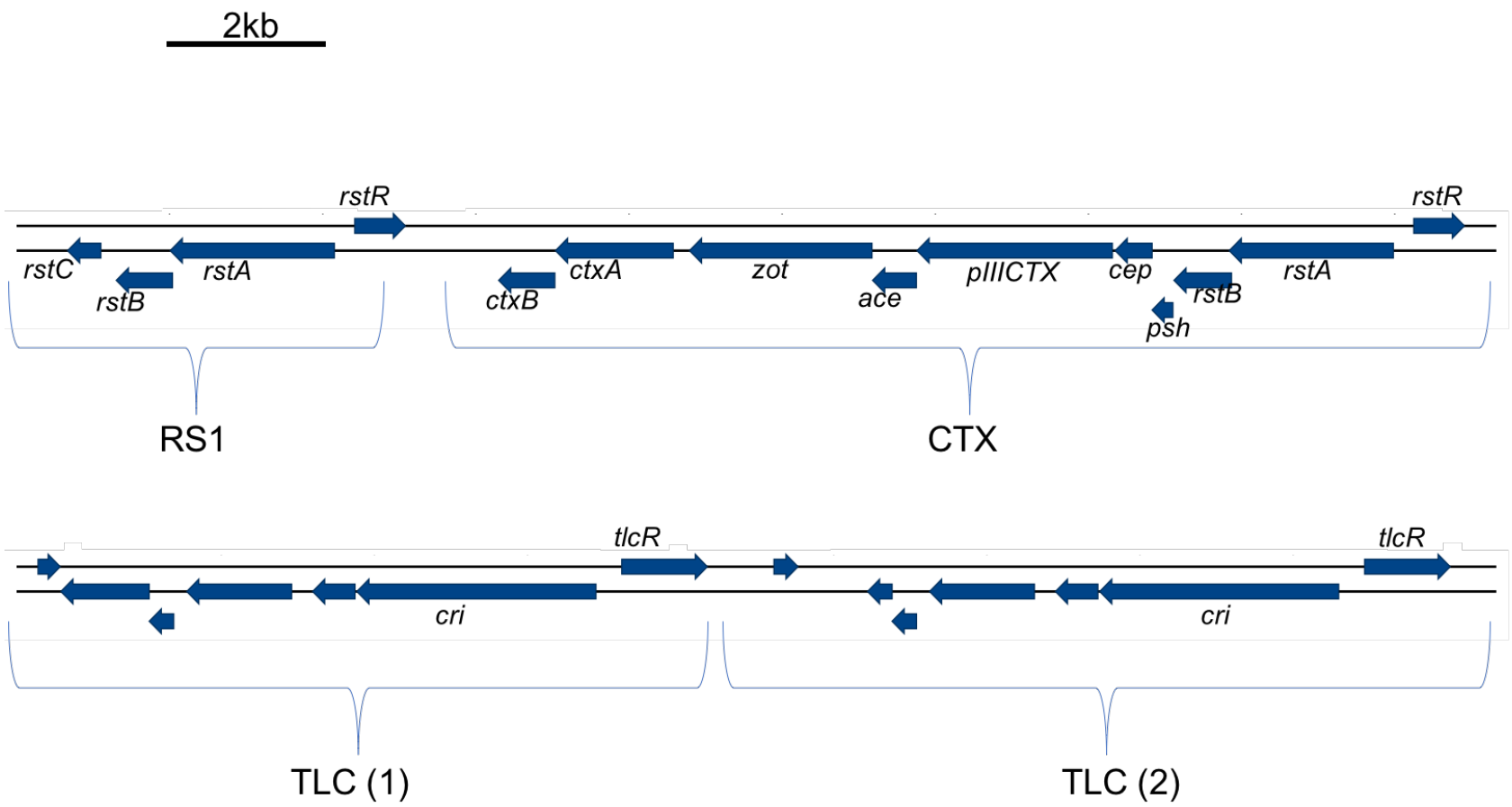
B



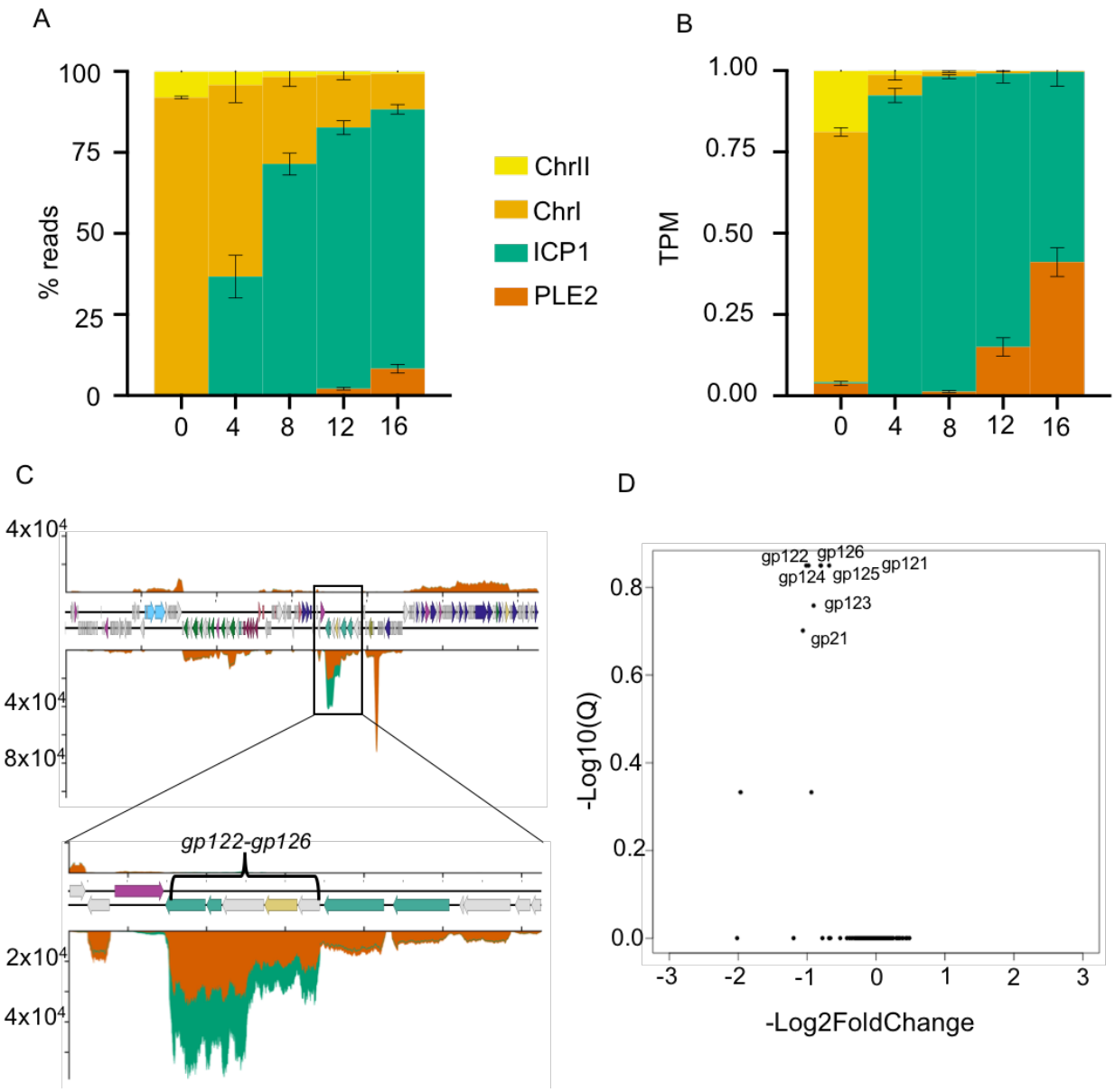
Supplemental Fig 2.10 Conservation of the PLE IIR transcript (**A**) Clustal alignment of the PLE IIR transcript across all PLEs. Terminal inverted repeats are highlighted in grey. Sequence with antisense homology to PLE gene leader sequences is bolded. (**B**) Alignment of the an 'ideal' leader sequence antisense to the IIR, and the leader sequences of several genes in PLE1. Matching sites are bolded. Mismatches are shown in light gray.



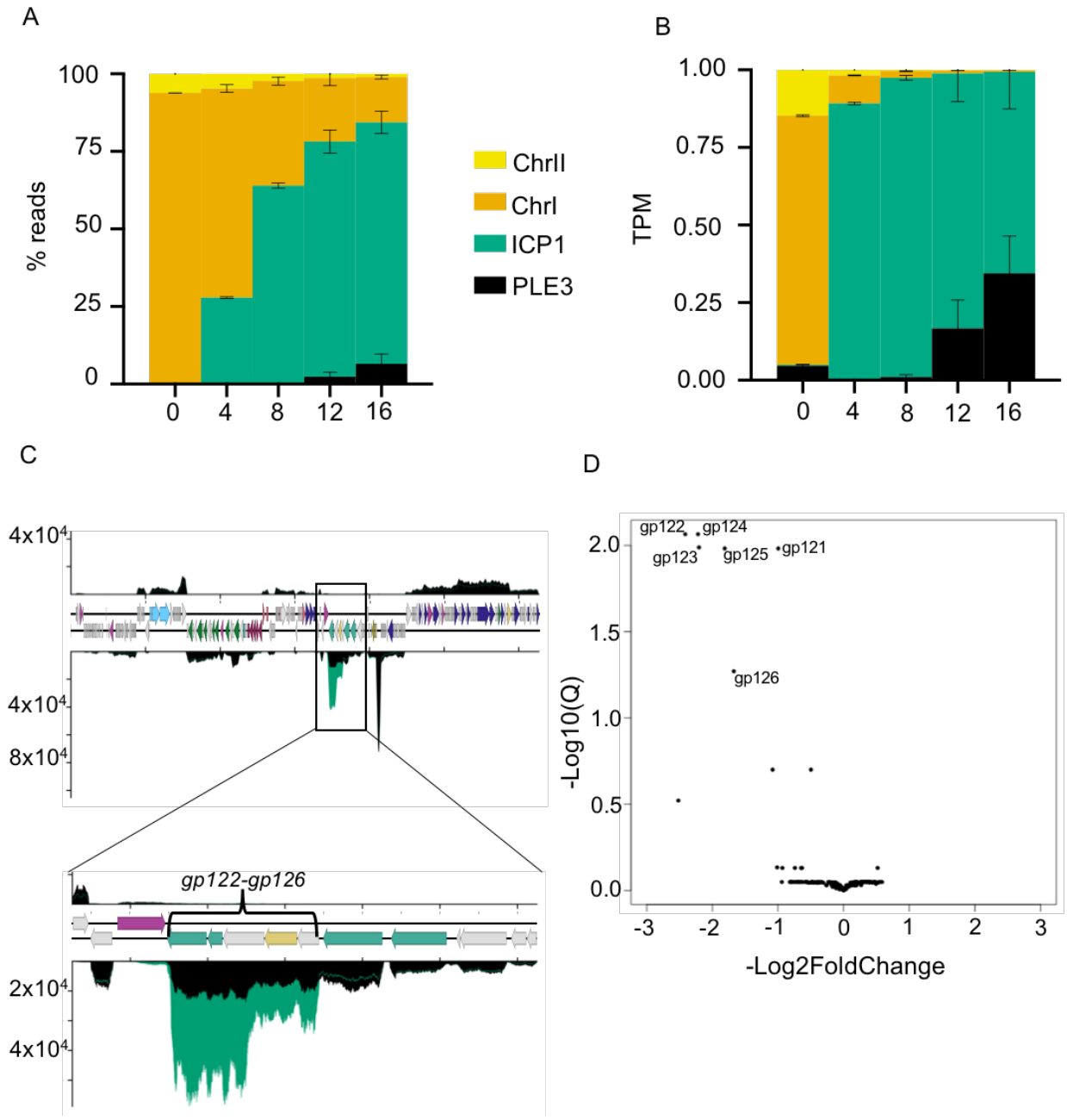
Supplemental Figure 2.11 Some *V. cholerae* genes are differentially expressed in the presence of PLEs Volcano plots showing differential regulation of *V. cholerae* genes in uninfected PLE(+) cultures relative to PLE(-) cultures. Genes within 2kb of the PLE integration site are colored red, other genes within the superintegron are colored yellow, and genes encoded in RS1,CTXφ, or TLC are colored blue. Genes that are not significantly differentially regulated are colored light gray, and other significantly differentially regulated genes are colored black. A Q value cut off of 0.1 was used to determine significance. PLEs 2 (A), 3 (B), 4 (C), and 5 (D) are shown.



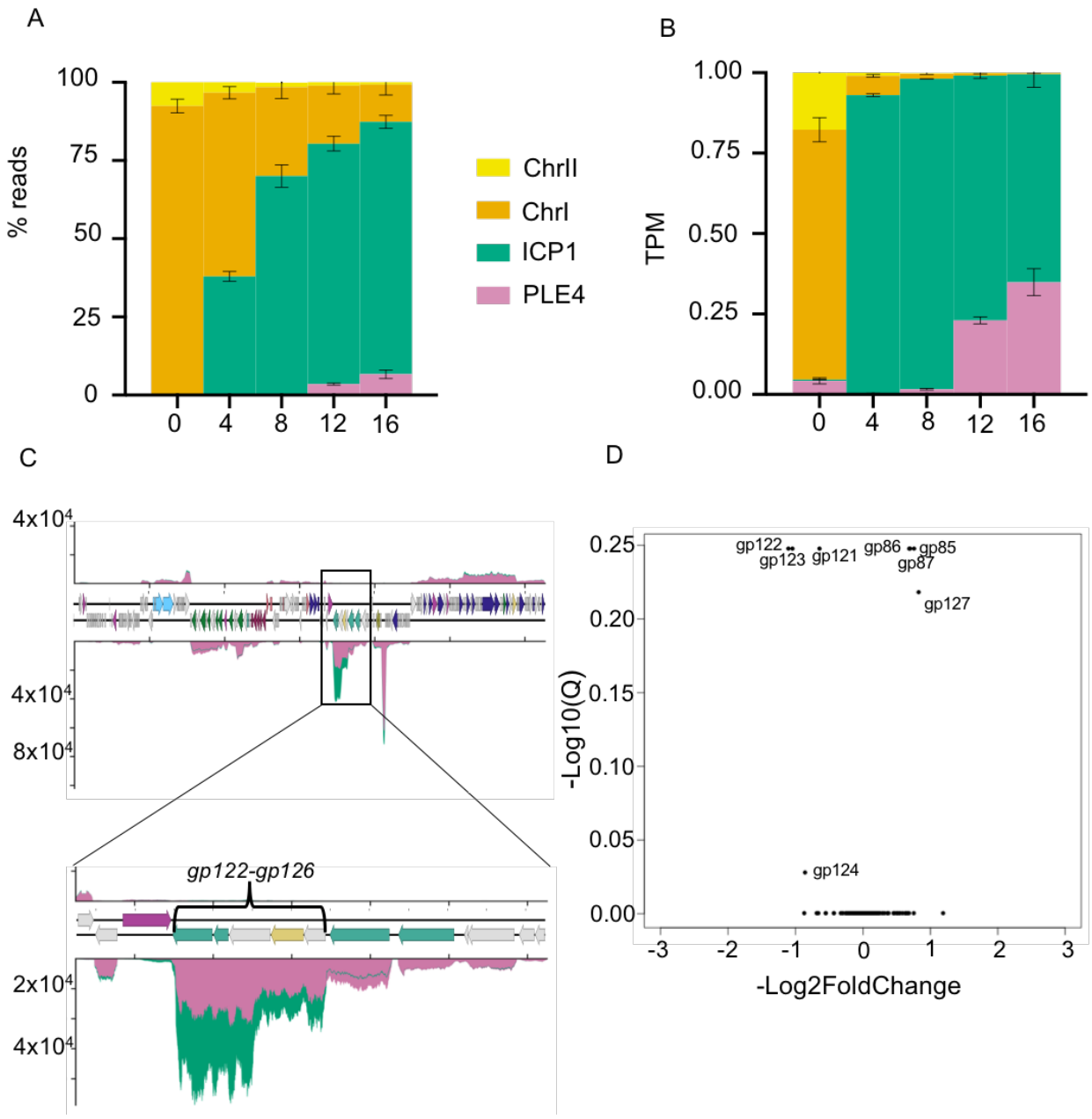
Supplemental Figure 2.12. Integrative mobile elements exploiting Xer in E7946 *V. cholerae* Diagram showing IMEX elements integrated at E7946's dif site on the large chromosome. Brackets denote mobile elements.



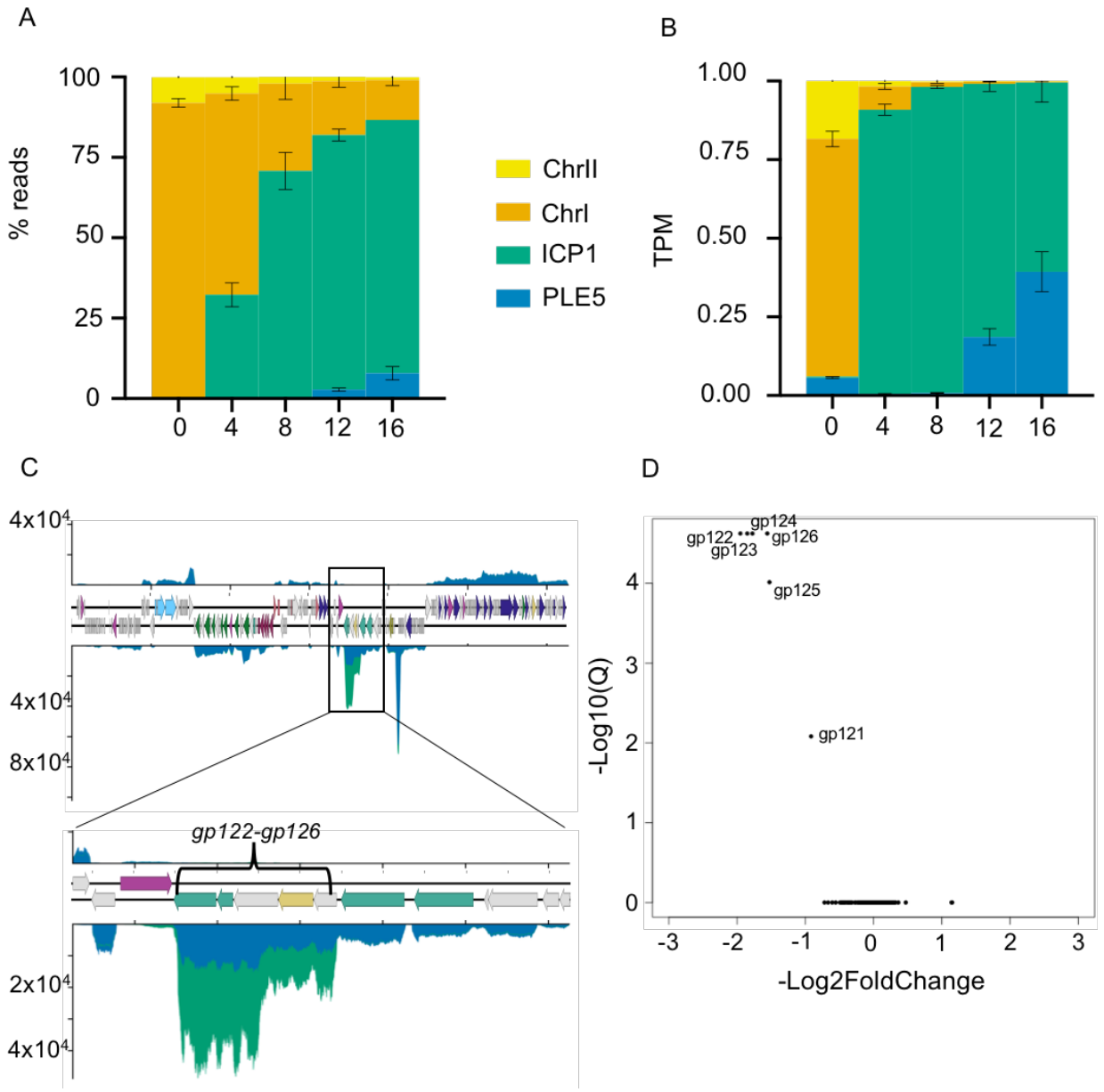
Supplemental Figure 2.13 PLE2 specifically downregulates ICP1 capsid production (**A**) Percent reads abundance for PLE2, *V. cholerae* chromosomes and ICP1 over the infection time course. (**B**) Reads normalized to a TPM value based on the total number of reads from each element, and the element's length. (**C**) Volcano plot of ICP1 differential gene expression in the PLE2 culture relative to the PLE(-) culture at 16 minutes post infection. (**D**) Average relative coverage along ICP1's genome in PLE(-) (green) and PLE2 (orange) cultures at 16 minutes post infection. The insert depicts ICP1's head morphogenesis operon. For all panels, results incorporate values obtained for two biological replicates.



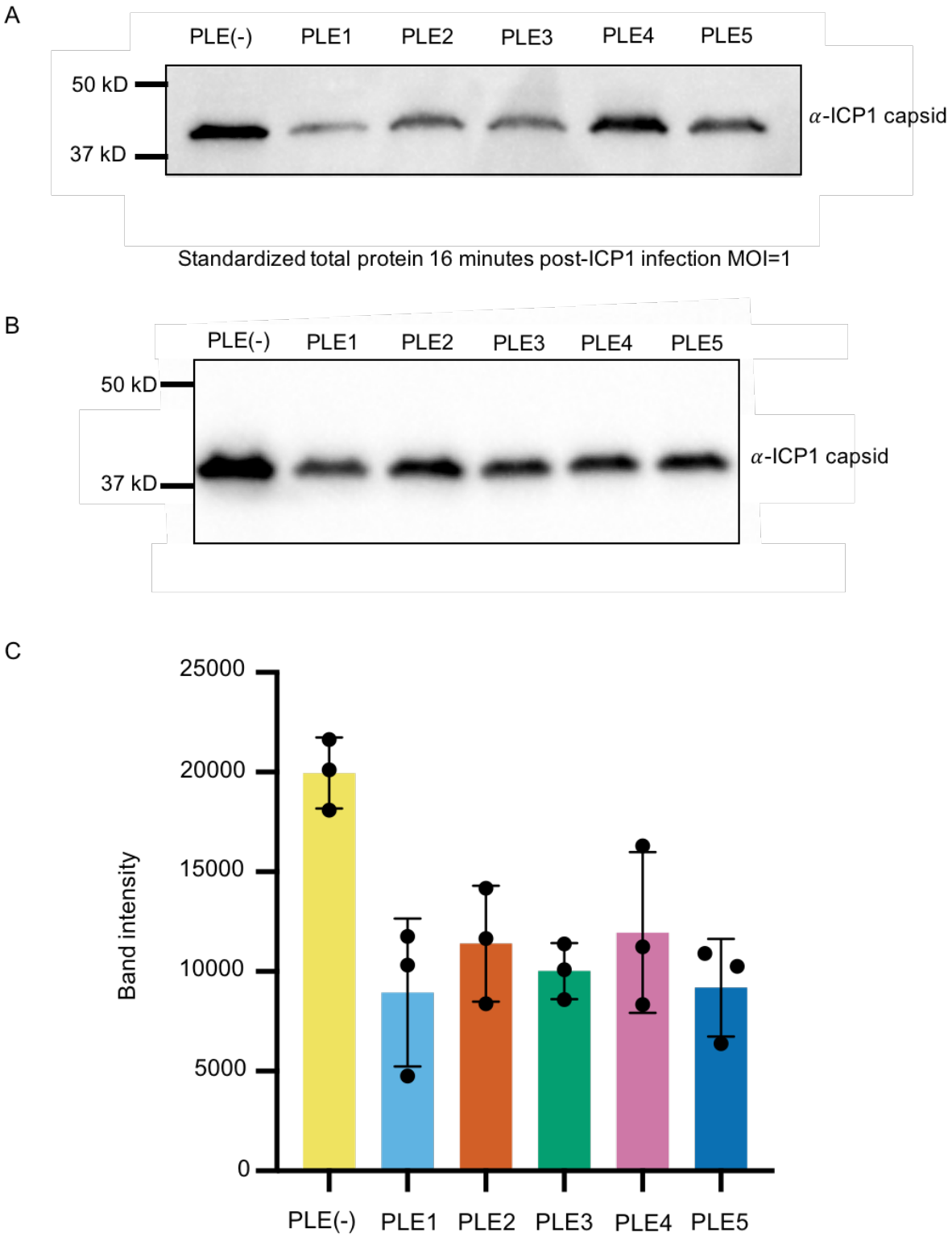
Supplemental Figure 2.14 PLE3 specifically downregulates ICP1 capsid production (**A**) Percent reads abundance for PLE3, *V. cholerae* chromosomes and ICP1 over the infection time course. (**B**) Reads normalized to a TPM value based on the total number of reads from each element, and the element's length. (**C**) Volcano plot of ICP1 differential gene expression in the PLE3 culture relative to the PLE(-) culture at 16 minutes post infection. (**D**) Average relative coverage along ICP1's genome in PLE(-) (green) and PLE3 (black) cultures at 16 minutes post infection. The insert depicts ICP1's head morphogenesis operon. For all panels, results incorporate values obtained for two biological replicates.



Supplemental Figure 2.15 PLE4 specifically downregulates ICP1 capsid production (**A**) Percent reads abundance for PLE4, *V. cholerae* chromosomes and ICP1 over the infection time course. (**B**) Reads normalized to a TPM value based on the total number of reads from each element, and the element's length. (**C**) Volcano plot of ICP1 differential gene expression in the PLE4 culture relative to the PLE(-) culture at 16 minutes post infection. (**D**) Average relative coverage along ICP1's genome in PLE(-) (green) and PLE4 (pink) cultures at 16 minutes post infection. The insert depicts ICP1's head morphogenesis operon. For all panels, results incorporate values obtained for two biological replicates



Supplemental Figure 2.16 PLE5 specifically downregulates ICP1 capsid production(**A**) Percent reads abundance for PLE5, *V. cholerae* chromosomes and ICP1 over the infection time course. (**B**) Reads normalized to a TPM value based on the total number of reads from each element, and the element's length. (**C**) Volcano plot of ICP1 differential gene expression in the PLE5 culture relative to the PLE(-) culture at 16 minutes post infection. (**D**) Average relative coverage along ICP1's genome in PLE(-) (green) and PLE5 (blue) cultures at 16 minutes post infection. The insert depicts ICP1's head morphogenesis operon. For all panels, results incorporate values obtained for two biological replicates.



Supplemental Figure 2.17 PLEs reduce the level of expressed capsid protein (A) and **(B)** Biological replicates of Western blots against Gp122, ICP1's major capsid protein from infections of PLE(-), PLE1, PLE2, PLE3, PLE4, and PLE5 cultures at 16 minutes post infection. **(C)** Quantification of the Western blot band intensities from (A) (B) and Fig 7E.

Supplementary Table S1.1: Strains used in this study

Strain	Description*	Source
KDS6	*V. cholerae* O1, El Tor biotype	Lab collection
KDS36	*V. cholerae* E7946 containing PLE1	[29]
KDS153	*V. cholerae* E7946 PLE1 ΔORF7::frt, KanR	This study
KDS154	*V. cholerae* E7946 PLE1 ΔORF8::frt, KanR	This study
KDS155	*V. cholerae* E7946 PLE1 ΔORF9::frt, KanR	This study
KDS156	*V. cholerae* E7946 PLE1 ΔORF10::frt, KanR	This study
KDS157	*V. cholerae* E7946 PLE1 *ΔrepA*::frt, KanR	This study
KDS158	*V. cholerae* E7946 PLE1 ΔORF12::frt, KanR	This study
KDS159	*V. cholerae* E7946 PLE1 ΔORF12.1::frt, KanR	This study
KDS160	*V. cholerae* E7946 PLE1 ΔORF13::frt, KanR	This study
KDS161	*V. cholerae* E7946 PLE1 ΔORF14::frt, KanR	This study
KDS181	*V. cholerae* E7946 PLE1 *Δint*::Spec-frt	[55]
KDS182	*V. cholerae* E7946 PLE1 ΔORFs2-5::Spec-frt	[55]
KDS183	*V. cholerae* E7946 PLE1 ΔORFs7-14::Spec-frt	[55]
KDS184	*V. cholerae* E7946 PLE1 ΔORFs15-20::Spec-frt	[55]
KDS185	*V. cholerae* E7946 PLE1 ΔORFs21-23::Spec-frt	[55]
KDS228	*V. cholerae* E7946 *ΔlacZ*::SpecR	[29]
KDS229	*V. cholerae* E7946 PLE1 *ΔrepA*::frt, *ΔlacZ*::P_{tac}-*repA*, KanR, SpecR (RepA chromosomal expression construct in PLE *ΔrepA*)	This study
KDS230	*V. cholerae* E7946 PLE1 *ΔrepA*::frt, *ΔlacZ*:: P_{tac} *EV*, KanR, SpecR (Empty chromosomal expression construct in PLE *ΔrepA*)	This study
KDS231	*V. cholerae* E7946 PLE1 *ΔrepA*::frt, P_{tac}-*repA*, KanR, CmR (Plasmid RepA expression construct)	This study
KDS232	*V. cholerae* E7946 midiPLE, P_{tac}-*repA*, KanR, SpecR (RepA plasmid expression construct in strain with midiPLE)	This study
KDS233	*V. cholerae* E7946 PLE1 *ΔrepA*::frt, P_{tac} *EV*, KanR, CmR (Empty plasmid expression construct in PLE *ΔrepA*)	This study
KDS234	*V. cholerae* E7946 midiPLE, P_{tac} *EV*, KanR, CmR (Empty plasmid expression construct in strain with midiPLE)	This study
KDS235	*V. cholerae* E7946 PLE1 ΔNCR1, KanR	This study
KDS236	*V. cholerae* E7946 PLE1 ΔNCR2::frt, KanR	This study
KDS237	*V. cholerae* E7946 PLE1 ΔNCR3::frt, KanR	This study
KDS238	*E. coli* BL21 pE-SUMO-RepA. Vector to express 6xHisSumo-fusion protein, fused to N-terminus of RepA	This study
KDS263	*V. cholerae* E7946 PLE1 ∆repeat 3::frt	This study
KDS264	*V. cholerae* E7946 PLE1 ∆repeat 4::frt	This study
KDS265	*V. cholerae* E7946 PLE1 ∆repeat 3::frt, *ΔrepA*::frt, KanR	

Strain	Description*	Source
ICP1	ICP1_2006_E ΔCRISPR ΔCas2-3	[55]

* KanR = Kanamycin resistance cassette, SpecR = Spectinomycin resistance cassette.

Supplementary Table S2: Primers used in this study

Primer	Sequence	Purpose	Source
zac14	AGGGTTTGAGTGCGATTACG	qPCR PLE	[29]
zac15	TGAGGTTTTACCACCTTTTGC	qPCR PLE	[29]
zac68	CTGAATCGCCCTACCCGTAC	qPCR ICP1	[29]
zac69	GTGAACCAACCTTTGTCGCC	qPCR ICP1	[29]
zac109	CGCCAAACCAACAAGACAGG	qPCR PLE ΔNCR	This study
zac110	CCCCAAGATCAACCACCTCC	qPCR PLE ΔNCR	This study
zac205	TTTATATGGATGATTTTCAACCCTATAAA	F sequence R3 probe	This study
zac206	TTTATAGGGTTGAAAATCATCCATATAAA	R sequence R3 probe	This study
zac207	AAGGTAGACACCTTATGGTAGTTC	F primer R4 probe	This study
zac208	GAACTACCATAAGGTGTCTACCTT	R primer R4 probe	This study

Supplementary Table S1.3: Proportional reads abundance for ICP1, PLE, and the *V. cholerae* chromosomes relative to total sample reads over time course of ICP1 infection

Percent reads per element of total reads per condition							
	PLE(-) infection			PLE(+) infection			
Min[a]	VC I[b]	VC II	ICP1	VC I	VC II	ICP1	PLE
4	79 ± 0. 27[c]	20 ± 0.098	1.1 ± 0.17	79 ± 0.26	20 ± 0.15	1.1 ± 0.26	0. 32 ± 8.6E-04
8	78 ± 6.0	20 ± 0.097	2.2 ± 0.097	77 ± 0.57	19 ± 0.14	2.5 ± 0.61	0.67 ± 0.12
12	61 ± 5.9	15 ± 1.6	24 ± 7.4	65 ± 0.71	17 ± 0.17	14 ± 0.71	4.5 ± 0.29
16	41 ± 4.8	9.9 ± 1.2	49 ± 6.0	51 ± 1.3	14 ± 0.35	17 ± 0.69	19 ± 1.4

[a] time in minutes post-infection
[b] VC I is the *V. cholerae* large chromosome, VCI II is the *V. cholerae* small chromosome
[c] Values are the mean and standard deviation of three biological replicates

Supplementary Table S1.4: Proportional reads abundance for ICP1, PLE, and the *V. cholerae* chromosomes relative to total sample reads over time course of ICP1 infection in PLE *ΔrepA*

Percent reads per element of total reads per condition				
Min[a]	VC I	VC II	ICP1	PLE
8	77 ± 0.56	19 ± 0.12	30 ± 0.62	0.31 ± 0.013
16	44 ± 3.1	11 ± 0.8	45 ± 0.39	0.24 ± 8.5E -03

[a] time in minutes post-infection
[b] VC I is the *V. cholerae* large chromosome, VCI II is the *V. cholerae* small chromosome
[c] Values are the mean and standard deviation of three biological replicates

Supplementary Table S2.1: Strains used in this study

Strain	Description*	Source
KDS6	*V. cholerae* E7946 O1, El Tor biotype	Lab collection
KDS36	*V. cholerae* E7946 containing PLE1	[29]
KDS37	*V. cholerae* E7946 containing PLE2	[29]
KDS38	*V. cholerae* E7946 containing PLE3	[29]
KDS39	*V. cholerae* E7946 containing PLE4	[29]
KDS40	*V. cholerae* E7946 containing PLE5	[29]
KDS281	*V. cholerae* E7946 Δ*ctxAB*::KanR	This study
KDS282	*V. cholerae* E7946 Δ*ctxAB*::KanR containing PLE1	This study
KDS283	*V. cholerae* E7946 Δ*ctxAB*::KanR containing PLE2	This study
KDS284	*V. cholerae* E7946 ΔCTXϕ, Δlacz::P_{tac}-*toxT,* SpecR	This study
ICP1	ICP1_2006_E ΔCRISPR ΔCas2-3	[55]

* KanR = Kanamycin resistance cassette, SpecR = Spectinomycin resistance cassette.

References

1. Slobedman B, Barry PA, Spencer JV, Avdic S, Abendroth A. Virus-encoded homologs of cellular interleukin-10 and their control of host immune function. J Virol. 2009;83: 9618–9629. doi:10.1128/JVI.01098-09
2. Suthaus J, Adam N, Grötzinger J, Scheller J, Rose-John S. Viral Interleukin-6: Structure, pathophysiology and strategies of neutralization. Eur J Cell Biol. 2011;90: 495–504. doi:10.1016/j.ejcb.2010.10.016
3. Andrietti F, Polidori C. The Hidden Biodiversity Data Retained in Pre-Linnaean Works: A Case Study with Two Important XVII Century Italian Entomologists. In: Casetta E, Marques da Silva J, Vecchi D, editors. From assessing to conserving biodiversity: conceptual and practical challenges. Cham: Springer International Publishing; 2019. pp. 21–54. doi:10.1007/978-3-030-10991-2_2
4. Harvey JA, Poelman EH, Tanaka T. Intrinsic inter- and intraspecific competition in parasitoid wasps. Annu Rev Entomol. 2013;58: 333–351. doi:10.1146/annurev-ento-120811-153622
5. van Lenteren JC, Godfray HCJ. European science in the Enlightenment and the discovery of the insect parasitoid life cycle in The Netherlands and Great Britain. Biological Control. 2005;32: 12–24. doi:10.1016/j.biocontrol.2004.08.009
6. Darwin Correspondence Project, “Letter no. 2814” [Internet]. [cited 9 Apr 2020]. Available: https://www.darwinproject.ac.uk/letter/DCP-LETT-2814.xml
7. Darwin C. On the Origin of Species: Or; The Preservation of the Favoured Races in the Struggle for Life. Read Books Ltd; 2018.
8. Forbes AA, Bagley RK, Beer MA, Hippee AC, Widmayer HA. Quantifying the unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal order. BMC Ecol. 2018;18: 21. doi:10.1186/s12898-018-0176-x
9. Askew RR. On the biology of the inhabitants of oak galls of Cynipidae (Hymenoptera) in Britain. TRANSACTIONS OF THE SOCIETY FOR BRITISH ENTOMOLOGY. 1961;14: 237–268.
10. Georgiades K, Raoult D. Defining pathogenic bacterial species in the genomic era. Front Microbiol. 2010;1: 151. doi:10.3389/fmicb.2010.00151
11. Niu C, Yu D, Wang Y, Ren H, Jin Y, Zhou W, et al. Common and pathogen-specific virulence factors are different in function and structure. Virulence. 2013;4: 473–482. doi:10.4161/viru.25730
12. Brown SP, Cornforth DM, Mideo N. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. Trends Microbiol. 2012;20: 336–342. doi:10.1016/j.tim.2012.04.005
13. Ana A, Regina C W LaRocque. 47 - Cholera and Other Vibrios. In: Edward T, David R, Tom, Naomi E, Timothy, P R Hill, Solomon, Aronson, Endy, editor. Hunter’s Tropical Medicine and Emerging Infectious Diseases. 10th ed. Content Repository Only!; 2020. pp. 486–491.
14. Clokie MR, Millard AD, Letarov AV, Heaphy S. Phages in nature. Bacteriophage. 2011;1: 31–45. doi:10.4161/bact.1.1.14942
15. Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol. 1952;36: 39–56. doi:10.1085/jgp.36.1.39
16. Benzer S. Fine structure of a genetic region in bacteriophage. Proc Natl Acad Sci USA. 1955;41: 344–354.

17. Benzer S. On the topology of the genetic fine structure. Proc Natl Acad Sci USA. 1959;45: 1607–1620.
18. Crick FH, Barnett L, Brenner S, Watts-Tobin RJ. General nature of the genetic code for proteins. Nature. 1961;192: 1227–1232. doi:10.1038/1921227a0
19. Grosman AH, Janssen A, de Brito EF, Cordeiro EG, Colares F, Fonseca JO, et al. Parasitoid increases survival of its pupae by inducing hosts to fight predators. PLoS One. 2008;3: e2276. doi:10.1371/journal.pone.0002276
20. Maure F, Brodeur J, Ponlet N, Doyon J, Firlej A, Elguero E, et al. The cost of a bodyguard. Biol Lett. 2011;7: 843–846. doi:10.1098/rsbl.2011.0415
21. Thomas F, Schmidt-Rhaesa A, Martin G, Manu C, Durand P, Renaud F. Do hairworms (Nematomorpha) manipulate the water seeking behaviour of their terrestrial hosts? J Evol Biol. 2002;15: 356–361. doi:10.1046/j.1420-9101.2002.00410.x
22. Berdoy M, Webster JP, Macdonald DW. Fatal attraction in rats infected with Toxoplasma gondii. Proc Biol Sci. 2000;267: 1591–1594. doi:10.1098/rspb.2000.1182
23. Boyce GR, Gluck-Thaler E, Slot JC, Stajich JE, Davis WJ, James TY, et al. Psychoactive plant- and mushroom-associated alkaloids from two behavior modifying cicada pathogens. Fungal Ecol. 2019;41: 147–164. doi:10.1016/j.funeco.2019.06.002
24. Loenen WAM, Dryden DTF, Raleigh EA, Wilson GG, Murray NE. Highlights of the DNA cutters: a short history of the restriction enzymes. Nucleic Acids Res. 2014;42: 3–19. doi:10.1093/nar/gkt990
25. Pickar-Oliver A, Gersbach CA. The next generation of CRISPR-Cas technologies and applications. Nat Rev Mol Cell Biol. 2019;20: 490–507. doi:10.1038/s41580-019-0131-5
26. Seed KD, Bodi KL, Kropinski AM, Ackermann H-W, Calderwood SB, Qadri F, 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. 2011;2: e00334-10. doi:10.1128/mBio.00334-10
27. Seed KD, Faruque SM, Mekalanos JJ, Calderwood SB, Qadri F, Camilli A. Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in Vibrio cholerae O1. PLoS Pathog. 2012;8: e1002917. doi:10.1371/journal.ppat.1002917
28. Silva-Valenzuela CA, Camilli A. Niche adaptation limits bacteriophage predation of Vibrio cholerae in a nutrient-poor aquatic environment. Proc Natl Acad Sci USA. 2019;116: 1627–1632. doi:10.1073/pnas.1810138116
29. O'Hara BJ, Barth ZK, McKitterick AC, Seed KD. A highly specific phage defense system is a conserved feature of the Vibrio cholerae mobilome. PLoS Genet. 2017;13: e1006838. doi:10.1371/journal.pgen.1006838
30. Barth ZK, Silvas TV, Angermeyer A, Seed KD. Genome replication dynamics of a bacteriophage and its satellite reveal strategies for parasitism and viral restriction. Nucleic Acids Res. 2020;48: 249–263. doi:10.1093/nar/gkz1005
31. Krupovic M, Kuhn JH, Fischer MG. A classification system for virophages and satellite viruses. Arch Virol. 2016;161: 233–247. doi:10.1007/s00705-015-2622-9
32. Christie GE, Dokland T. Pirates of the Caudovirales. Virology. 2012;434: 210–221. doi:10.1016/j.virol.2012.10.028
33. Contursi P, Fusco S, Cannio R, She Q. Molecular biology of fuselloviruses and their satellites. Extremophiles. 2014;18: 473–489. doi:10.1007/s00792-014-0634-0
34. Hughes SA, Wedemeyer H, Harrison PM. Hepatitis delta virus. Lancet. 2011;378: 73–85.

doi:10.1016/S0140-6736(10)61931-9
35. Simon AE, Roossinck MJ, Havelda Z. Plant virus satellite and defective interfering RNAs: new paradigms for a new century. Annu Rev Phytopathol. 2004;42: 415–437. doi:10.1146/annurev.phyto.42.040803.140402
36. Lindsay JA, Ruzin A, Ross HF, Kurepina N, Novick RP. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in Staphylococcus aureus. Mol Microbiol. 1998;29: 527–543. doi:10.1046/j.1365-2958.1998.00947.x
37. La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, et al. The virophage as a unique parasite of the giant mimivirus. Nature. 2008;455: 100–104. doi:10.1038/nature07218
38. Seed KD, Lazinski DW, Calderwood SB, Camilli A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature. 2013;494: 489–491. doi:10.1038/nature11927
39. Lindqvist BH, Dehò G, Calendar R. Mechanisms of genome propagation and helper exploitation by satellite phage P4. Microbiol Rev. 1993;57: 683–702.
40. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of Escherichia coli K-12. Science. 1997;277: 1453–1462. doi:10.1126/science.277.5331.1453
41. Penadés JR, Christie GE. The Phage-Inducible Chromosomal Islands: A Family of Highly Evolved Molecular Parasites. Annu Rev Virol. 2015;2: 181–201. doi:10.1146/annurev-virology-031413-085446
42. Fillol-Salom A, Martínez-Rubio R, Abdulrahman RF, Chen J, Davies R, Penadés JR. Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. ISME J. 2018;12: 2114–2128. doi:10.1038/s41396-018-0156-3
43. Diana C, Dehò G, Geisselsoder J, Tinelli L, Goldstein R. Viral interference at the level of capsid size determination by satellite phage P4. J Mol Biol. 1978;126: 433–445. doi:10.1016/0022-2836(78)90050-5
44. Ram G, Chen J, Kumar K, Ross HF, Ubeda C, Damle PK, et al. Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. Proc Natl Acad Sci USA. 2012;109: 16300–16305. doi:10.1073/pnas.1204615109
45. Ubeda C, Maiques E, Knecht E, Lasa I, Novick RP, Penadés JR. Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. Mol Microbiol. 2005;56: 836–844. doi:10.1111/j.1365-2958.2005.04584.x
46. Christie GE, Calendar R. Bacteriophage P2. Bacteriophage. 2016;6: e1145782. doi:10.1080/21597081.2016.1145782
47. Christie GE, Matthews AM, King DG, Lane KD, Olivarez NP, Tallent SM, et al. The complete genomes of Staphylococcus aureus bacteriophages 80 and 80α--implications for the specificity of SaPI mobilization. Virology. 2010;407: 381–390. doi:10.1016/j.virol.2010.08.036
48. Hood IV, Berger JM. Viral hijacking of a replicative helicase loader and its implications for helicase loading control and phage replication. Elife. 2016;5. doi:10.7554/eLife.14158
49. Ziegelin G, Lanka E. Bacteriophage P4 DNA replication. FEMS Microbiol Rev. 1995;17: 99–107. doi:10.1111/j.1574-6976.1995.tb00191.x
50. Ubeda C, Barry P, Penadés JR, Novick RP. A pathogenicity island replicon in

Staphylococcus aureus replicates as an unstable plasmid. Proc Natl Acad Sci USA. 2007;104: 14182–14188. doi:10.1073/pnas.0705994104
51. Ubeda C, Tormo-Más MÁ, Penadés JR, Novick RP. Structure-function analysis of the SaPIbov1 replication origin in Staphylococcus aureus. Plasmid. 2012;67: 183–190. doi:10.1016/j.plasmid.2012.01.006
52. Angermeyer A, Das MM, Singh DV, Seed KD. Analysis of 19 Highly Conserved Vibrio cholerae Bacteriophages Isolated from Environmental and Patient Sources Over a Twelve-Year Period. Viruses. 2018;10. doi:10.3390/v10060299
53. Kulczyk AW, Richardson CC. The replication system of bacteriophage T7. Enzymes. 2016;39: 89–136. doi:10.1016/bs.enz.2016.02.001
54. Weigel C, Seitz H. Bacteriophage replication modules. FEMS Microbiol Rev. 2006;30: 321–381. doi:10.1111/j.1574-6976.2006.00015.x
55. McKitterick AC, Seed KD. Anti-phage islands force their target phage to directly mediate island excision and spread. Nat Commun. 2018;9: 2348. doi:10.1038/s41467-018-04786-5
56. McKitterick AC, LeGault KN, Angermeyer A, Alam M, Seed KD. Competition between mobile genetic elements drives optimization of a phage-encoded CRISPR-Cas system: insights from a natural arms race. Philos Trans R Soc Lond B, Biol Sci. 2019;374: 20180089. doi:10.1098/rstb.2018.0089
57. Dalia AB, Lazinski DW, Camilli A. Identification of a membrane-bound transcriptional regulator that links chitin and natural competence in Vibrio cholerae. MBio. 2014;5: e01028-13. doi:10.1128/mBio.01028-13
58. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357–359. doi:10.1038/nmeth.1923
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079. doi:10.1093/bioinformatics/btp352
60. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol. 2014;1151: 165–188. doi:10.1007/978-1-4939-0554-6_12
61. Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Eng. 2007;9: 90–95. doi:10.1109/MCSE.2007.55
62. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 2012;131: 281–285. doi:10.1007/s12064-012-0162-3
63. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng. 1998;11: 739–747.
64. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. Protein Sci. 2018;27: 112–128. doi:10.1002/pro.3280
65. Rasmussen T, Jensen RB, Skovgaard O. The two chromosomes of Vibrio cholerae are initiated at different time points in the cell cycle. EMBO J. 2007;26: 3124–3131. doi:10.1038/sj.emboj.7601747
66. Ramachandran R, Ciaccia PN, Filsuf TA, Jha JK, Chattoraj DK. Chromosome 1 licenses chromosome 2 replication in Vibrio cholerae by doubling the crtS gene dosage. PLoS Genet. 2018;14: e1007426. doi:10.1371/journal.pgen.1007426

67. Olm MR, Brown CT, Brooks B, Firek B, Baker R, Burstein D, et al. Identical bacterial populations colonize premature infant gut, skin, and oral microbiomes and exhibit different in situ growth rates. Genome Res. 2017;27: 601–612. doi:10.1101/gr.213256.116
68. Garneau JR, Depardieu F, Fortier L-C, Bikard D, Monot M. PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. Sci Rep. 2017;7: 8292. doi:10.1038/s41598-017-07910-5
69. Rajewska M, Wegrzyn K, Konieczny I. AT-rich region and repeated sequences - the essential elements of replication origins of bacterial replicons. FEMS Microbiol Rev. 2012;36: 408–434. doi:10.1111/j.1574-6976.2011.00300.x
70. Rocha EPC. The replication-related organization of bacterial genomes. Microbiology (Reading, Engl). 2004;150: 1609–1627. doi:10.1099/mic.0.26974-0
71. Ubeda C, Maiques E, Barry P, Matthews A, Tormo MA, Lasa I, et al. SaPI mutations affecting replication and transfer and enabling autonomous replication in the absence of helper phage. Mol Microbiol. 2008;67: 493–503. doi:10.1111/j.1365-2958.2007.06027.x
72. Holm L. Benchmarking Fold Detection by DaliLite v.5. Bioinformatics. 2019; doi:10.1093/bioinformatics/btz536
73. Weaver KE, Kwong SM, Firth N, Francia MV. The RepA_N replicons of Gram-positive bacteria: a family of broadly distributed but narrow host range plasmids. Plasmid. 2009;61: 94–109. doi:10.1016/j.plasmid.2008.11.004
74. Kwong SM, Ramsay JP, Jensen SO, Firth N. Replication of staphylococcal resistance plasmids. Front Microbiol. 2017;8: 2279. doi:10.3389/fmicb.2017.02279
75. Schumacher MA, Tonthat NK, Kwong SM, Chinnam NB, Liu MA, Skurray RA, et al. Mechanism of staphylococcal multiresistance plasmid replication origin assembly by the RepA protein. Proc Natl Acad Sci USA. 2014;111: 9121–9126. doi:10.1073/pnas.1406065111
76. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol. 2018;430: 2237–2243. doi:10.1016/j.jmb.2017.12.007
77. Fujisawa H, Morita M. Phage DNA packaging. Genes Cells. 1997;2: 537–545. doi:10.1046/j.1365-2443.1997.1450343.x
78. McKitterick AC, Hays SG, Johura F-T, Alam M, Seed KD. Viral satellites exploit phage proteins to escape degradation of the bacterial host chromosome. Cell Host Microbe. 2019;26: 504–514.e4. doi:10.1016/j.chom.2019.09.006
79. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W. Bacteriophage T4 genome. Microbiol Mol Biol Rev. 2003;67: 86–156, table of contents.
80. Kreuzer KN, Brister JR. Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. Virol J. 2010;7: 358. doi:10.1186/1743-422X-7-358
81. Smith MCM. Phage-encoded Serine Integrases and Other Large Serine Recombinases. Microbiol Spectr. 2015;3. doi:10.1128/microbiolspec.MDNA3-0059-2014
82. Goh S, Hussain H, Chang BJ, Emmett W, Riley TV, Mullany P. Phage ϕC2 mediates transduction of Tn6215, encoding erythromycin resistance, between Clostridium difficile strains. MBio. 2013;4: e00840-13. doi:10.1128/mBio.00840-13
83. Barth ZK, Netter Z, Angermeyer A, Bhardwaj P, Seed K. A family of viral satellites manipulates invading virus gene expression and affects cholera toxin mobilization.

BioRxiv. 2020; doi:10.1101/2020.03.12.988147
84. Breitbart M, Delwart E, Rosario K, Segalés J, Varsani A, Ictv Report Consortium. ICTV virus taxonomy profile: circoviridae. J Gen Virol. 2017;98: 1997–1998. doi:10.1099/jgv.0.000871
85. Yutin N, Colson P, Raoult D, Koonin EV. Mimiviridae: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. Virol J. 2013;10: 106. doi:10.1186/1743-422X-10-106
86. McLeod SM, Kimsey HH, Davis BM, Waldor MK. CTXphi and Vibrio cholerae: exploring a newly recognized type of phage-host cell relationship. Mol Microbiol. 2005;57: 347–356. doi:10.1111/j.1365-2958.2005.04676.x
87. Pant A, Das B, Bhadra RK. CTX phage of Vibrio cholerae: Genomics and applications. Vaccine. 2019; doi:10.1016/j.vaccine.2019.06.034
88. Wahl MC, Sen R. Exploiting phage strategies to modulate bacterial transcription. Transcription. 2019;10: 222–230. doi:10.1080/21541264.2019.1684137
89. Ceyssens P-J, Minakhin L, Van den Bossche A, Yakunina M, Klimuk E, Blasdel B, et al. Development of giant bacteriophage ϕKZ is independent of the host transcription apparatus. J Virol. 2014;88: 10501–10510. doi:10.1128/JVI.01347-14
90. Leskinen K, Blasdel BG, Lavigne R, Skurnik M. RNA-Sequencing Reveals the Progression of Phage-Host Interactions between φR1-37 and Yersinia enterocolitica. Viruses. 2016;8: 111. doi:10.3390/v8040111
91. Chevallereau A, Blasdel BG, De Smet J, Monot M, Zimmermann M, Kogadeeva M, et al. Next-Generation "-omics" Approaches Reveal a Massive Alteration of Host RNA Metabolism during Bacteriophage Infection of Pseudomonas aeruginosa. PLoS Genet. 2016;12: e1006134. doi:10.1371/journal.pgen.1006134
92. Blasdel BG, Chevallereau A, Monot M, Lavigne R, Debarbieux L. Comparative transcriptomics analyses reveal the conservation of an ancestral infectious strategy in two bacteriophage genera. ISME J. 2017;11: 1988–1996. doi:10.1038/ismej.2017.63
93. Danis-Wlodarczyk K, Blasdel BG, Jang HB, Vandenheuvel D, Noben J-P, Drulis-Kawa Z, et al. Genomic, transcriptomic, and structural analysis of Pseudomonas virus PA5oct highlights the molecular complexity among Jumbo phages. BioRxiv. 2018; doi:10.1101/406421
94. Sacher JC, Flint A, Butcher J, Blasdel B, Reynolds HM, Lavigne R, et al. Transcriptomic Analysis of the Campylobacter jejuni Response to T4-Like Phage NCTC 12673 Infection. Viruses. 2018;10. doi:10.3390/v10060332
95. Blasdel BG, Ceyssens P-J, Chevallereau A, Debarbieux L, Lavigne R. Comparative transcriptomics reveals a conserved Bacterial Adaptive Phage Response (BAPR) to viral predation. BioRxiv. 2018; doi:10.1101/248849
96. Gaia M, Benamar S, Boughalmi M, Pagnier I, Croce O, Colson P, et al. Zamilon, a novel virophage with Mimiviridae host specificity. PLoS One. 2014;9: e94923. doi:10.1371/journal.pone.0094923
97. Hays SG, Seed KD. Dominant *Vibrio cholerae* phage exhibits lysis inhibition sensitive to disruption by a defensive phage satellite. BioRxiv. 2019; doi:10.1101/790493
98. Box AM, McGuffie MJ, O'Hara BJ, Seed KD. 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. 2016;198: 578–590. doi:10.1128/JB.00747-15

99. Dalia AB, McDonough E, Camilli A. Multiplex genome editing by natural transformation. Proc Natl Acad Sci USA. 2014;111: 8937–8942. doi:10.1073/pnas.1406478111
100. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550. doi:10.1186/s13059-014-0550-8
101. Hsiao A, Liu Z, Joelsson A, Zhu J. Vibrio cholerae virulence regulator-coordinated evasion of host immunity. Proc Natl Acad Sci USA. 2006;103: 14542–14547. doi:10.1073/pnas.0604650103
102. Utada AS, Bennett RR, Fong JCN, Gibiansky ML, Yildiz FH, Golestanian R, et al. Vibrio cholerae use pili and flagella synergistically to effect motility switching and conditional surface attachment. Nat Commun. 2014;5: 4913. doi:10.1038/ncomms5913
103. Conner JG, Zamorano-Sánchez D, Park JH, Sondermann H, Yildiz FH. The ins and outs of cyclic di-GMP signaling in Vibrio cholerae. Curr Opin Microbiol. 2017;36: 20–29. doi:10.1016/j.mib.2017.01.002
104. Davies BW, Bogard RW, Young TS, Mekalanos JJ. Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for V. cholerae virulence. Cell. 2012;149: 358–370. doi:10.1016/j.cell.2012.01.053
105. Cohen D, Melamed S, Millman A, Shulman G, Oppenheimer-Shaanan Y, Kacen A, et al. Cyclic GMP-AMP signalling protects bacteria against viral infection. Nature. 2019;574: 691–695. doi:10.1038/s41586-019-1605-5
106. Michael AJ. Biosynthesis of polyamines and polyamine-containing molecules. Biochem J. 2016;473: 2315–2329. doi:10.1042/BCJ20160185
107. Tabor CW, Tabor H. Polyamines in microorganisms. Microbiol Rev. 1985;49: 81–99.
108. Roos WH, Ivanovska IL, Evilevitch A, Wuite GJL. Viral capsids: mechanical characteristics, genome packaging and delivery mechanisms. Cell Mol Life Sci. 2007;64: 1484–1497. doi:10.1007/s00018-007-6451-1
109. Kahana C. Antizyme and antizyme inhibitor, a regulatory tango. Cell Mol Life Sci. 2009;66: 2479–2488. doi:10.1007/s00018-009-0033-3
110. Takatsuka Y, Yamaguchi Y, Ono M, Kamio Y. Gene cloning and molecular characterization of lysine decarboxylase from Selenomonas ruminantium delineate its evolutionary relationship to ornithine decarboxylases from eukaryotes. J Bacteriol. 2000;182: 6732–6741. doi:10.1128/jb.182.23.6732-6741.2000
111. Murata A, Odaka M, Mukuno S. The Bacteriophage-Inactivating Effect of Basic Amino Acids; Arginine, Histidine, and Lysine. Agric Biol Chem. 1974;38: 477–478. doi:10.1271/bbb1961.38.477
112. Watanabe T, Saito H. Inactivation of coliphage T2H by basic amino acids. Microbiol Immunol. 1978;22: 167–172. doi:10.1111/j.1348-0421.1978.tb00360.x
113. Kwiatkowski B, Beilharz H, Stirm S. Disruption of Vi bacteriophage III and localization of its deacetylase activity. J Gen Virol. 1975;29: 267–280. doi:10.1099/0022-1317-29-3-267
114. Stewart CR, Casjens SR, Cresawn SG, Houtz JM, Smith AL, Ford ME, et al. The genome of Bacillus subtilis bacteriophage SPO1. J Mol Biol. 2009;388: 48–70. doi:10.1016/j.jmb.2009.03.009
115. Kutter E, Bryan D, Ray G, Brewster E, Blasdel B, Guttman B. From Host to Phage Metabolism: Hot Tales of Phage T4's Takeover of E. coli. Viruses. 2018;10.

doi:10.3390/v10070387
116. Bentley SD, Parkhill J. Comparative genomic structure of prokaryotes. Annu Rev Genet. 2004;38: 771–792. doi:10.1146/annurev.genet.38.072902.094318
117. Merrikh H, Zhang Y, Grossman AD, Wang JD. Replication-transcription conflicts in bacteria. Nat Rev Microbiol. 2012;10: 449–458. doi:10.1038/nrmicro2800
118. Harris KA, Breaker RR. Large noncoding rnas in bacteria. Microbiol Spectr. 2018;6. doi:10.1128/microbiolspec.RWR-0005-2017
119. Delihas N. Small mobile sequences in bacteria display diverse structure/function motifs. Mol Microbiol. 2008;67: 475–481. doi:10.1111/j.1365-2958.2007.06068.x
120. Wagner EGH, Altuvia S, Romby P. Antisense RNAs in bacteria and their genetic elements. Adv Genet. 2002;46: 361–398.
121. Ravin NV, Svarchevsky AN, Dehò G. The anti-immunity system of phage-plasmid N15: identification of the antirepressor gene and its control by a small processed RNA. Mol Microbiol. 1999;34: 980–994. doi:10.1046/j.1365-2958.1999.01658.x
122. Sabbattini P, Six E, Zangrossi S, Briani F, Ghisotti D, Deho G. Immunity specificity determinants in the P4-like retronphage phi R73. Virology. 1996;216: 389–396. doi:10.1006/viro.1996.0074
123. Sherwood AV, Henkin TM. Riboswitch-Mediated Gene Regulation: Novel RNA Architectures Dictate Gene Expression Responses. Annu Rev Microbiol. 2016;70: 361–374. doi:10.1146/annurev-micro-091014-104306
124. Ayala JC, Silva AJ, Benitez JA. H-NS: an overarching regulator of the Vibrio cholerae life cycle. Res Microbiol. 2017;168: 16–25. doi:10.1016/j.resmic.2016.07.007
125. Ayala JC, Wang H, Silva AJ, Benitez JA. Repression by H-NS of genes required for the biosynthesis of the Vibrio cholerae biofilm matrix is modulated by the second messenger cyclic diguanylic acid. Mol Microbiol. 2015;97: 630–645. doi:10.1111/mmi.13058
126. Ayala JC, Wang H, Benitez JA, Silva AJ. RNA-Seq analysis and whole genome DNA-binding profile of the Vibrio cholerae histone-like nucleoid structuring protein (H-NS). Genom Data. 2015;5: 147–150. doi:10.1016/j.gdata.2015.05.039
127. Das B, Martínez E, Midonet C, Barre F-X. Integrative mobile elements exploiting Xer recombination. Trends Microbiol. 2013;21: 23–30. doi:10.1016/j.tim.2012.10.003
128. Dokland T. Scaffolding proteins and their role in viral assembly. Cell Mol Life Sci. 1999;56: 580–603. doi:10.1007/s000180050455
129. Birch EW, Ruggero NA, Covert MW. Determining host metabolic limitations on viral replication via integrated modeling and experimental perturbation. PLoS Comput Biol. 2012;8: e1002746. doi:10.1371/journal.pcbi.1002746
130. Mahmoudabadi G, Milo R, Phillips R. Energetic cost of building a virus. Proc Natl Acad Sci USA. 2017;114: E4324–E4333. doi:10.1073/pnas.1701670114
131. Linderoth NA, Calendar RL. The Psu protein of bacteriophage P4 is an antitermination factor for rho-dependent transcription termination. J Bacteriol. 1991;173: 6722–6731. doi:10.1128/jb.173.21.6722-6731.1991
132. Ranjan A, Banerjee R, Pani B, Sen U, Sen R. The moonlighting function of bacteriophage P4 capsid protein, Psu, as a transcription antiterminator. Bacteriophage. 2013;3: e25657. doi:10.4161/bact.25657
133. Ghosh G, Reddy J, Sambhare S, Sen R. A bacteriophage capsid protein is an inhibitor of a conserved transcription terminator of various bacterial pathogens. J Bacteriol. 2018;200. doi:10.1128/JB.00380-17

134. Lengyel JA, Calendar R. Control of bacteriophage P2 protein and DNA synthesis. Virology. 1974;57: 305–313. doi:10.1016/0042-6822(74)90170-6
135. Geisselsoder J, Chidambaram M, Goldstein R. Transcriptional control of capsid size in the P2:P4 bacteriophage system. J Mol Biol. 1978;126: 447–456. doi:10.1016/0022-2836(78)90051-7
136. Agarwal M, Arthur M, Arbeit RD, Goldstein R. Regulation of icosahedral virion capsid size by the in vivo activity of a cloned gene product. Proc Natl Acad Sci USA. 1990;87: 2428–2432. doi:10.1073/pnas.87.7.2428
137. Ram G, Chen J, Ross HF, Novick RP. Precisely modulated pathogenicity island interference with late phage gene transcription. Proc Natl Acad Sci USA. 2014;111: 14536–14541. doi:10.1073/pnas.1406749111
138. Mitarai N. How pirate phage interferes with helper phage: Comparison of the two distinct strategies. J Theor Biol. 2019; 110096. doi:10.1016/j.jtbi.2019.110096