

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Mobile genetic elements shape pathogen-phage coevolution in epidemic *Vibrio cholerae*

Permalink

<https://escholarship.org/uc/item/41p5p8n9>

Author

LeGault, Kristen

Publication Date

2022

Peer reviewed|Thesis/dissertation

Mobile genetic elements shape pathogen-phage coevolution in epidemic *Vibrio cholerae*

By

Kristen LeGault

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 Arash Komeili
Professor Kathleen Ryan

Spring 2022

© Kristen LeGault 2022

Abstract

Mobile genetic elements shape pathogen-phage coevolution in epidemic *Vibrio cholerae*

By

Kristen LeGault

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Kimberley Seed, Chair

All life contends with biotic stresses, primary among them are infection by viruses. Bacteria face the constant threat of infection by their own viruses, known as bacteriophages, or phages for short. To defend against phage infection, bacteria encode a remarkable diversity of anti-phage defense systems, constituting what can be considered a prokaryotic immune system. The explosion in sequenced bacterial genomes coupled with new bioinformatic approaches indicates that up to 10% of a bacterium's genome may encode defense systems. However, deciphering what systems are important for determining resistances in natural populations of bacteria and phages lags this remarkable discovery. The pathogen *Vibrio cholerae* is predated upon by a low diversity of lytic phages in both its aquatic environment and throughout the course of human infection. Through longitudinal sampling, we have uncovered the resistance determinants encoded by *V. cholerae* that determines its ability to thwart phages in nature. Here we describe how two families of phage defense systems govern *V. cholerae*'s resistance to the dominant lytic phages found in endemic regions. We show that these defense systems are part of *V. cholerae*'s mobilome, allowing for rapid adaptation to changing phage threats through acquisition of horizontally transferred defenses. Importantly, it is the pan-immune system of *V. cholerae* that allows it to keep pace with its lytic phages. We further describe phage isolates that can overcome co-circulating defenses, highlighting that ability for phages to overcome the dominant defense system, selecting for replacement of one cholera lineage with a new phage-resistant lineage. This work centers phage predation in the evolution and ecology of a globally important pathogen.

Table of Contents

Dedication	ii
Chapter 1: The phage-host arms race a driver of bacterial evolution	1
Chapter 2: Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts	5
Main Text	6
Discussion	18
Materials and Methods	21
Chapter 3: A phage parasite deploys a nicking nuclease effector to inhibit viral host replication	26
Introduction	27
Results	28
Discussion	41
Materials and Methods	44
Chapter 4: Toxicity of a PLE-encoded nuclease is mediated by a putative proteolytic adaptor protein	49
Introduction	50
Results	51
Discussion	56
Materials and Methods	58
Chapter 5: The future of phage biology: Centering phages in bacterial ecology and evolution	61
Supplementary Materials	65
References	97

Dedication

To my partner, adventure buddy, bestie and scientific hero Dr. Johan Jaenisch.
To my parents Anneli and Michael LeGault for their constant support.
To my sister Kath LeGault for keeping me happy through grad school.
To all the women scientists who paved the way for me,
especially my brilliant advisor Dr. Kim Seed

Chapter 1

The phage-host arms race a driver of bacterial evolution

All forms of life are plagued by the threat of being infected and killed by viruses. Viruses that infect bacteria, known as bacteriophages, are the most numerous viruses on the planet (1, 2), indicating that bacteria face constant assault from diverse and abundant phages. However, phages have not driven their hosts to extinction, illustrating the important role of phage immunity for bacterial fitness.

The bacterial membrane represents a first line of defense against phage infection. Phages co-opt host surface proteins for attachment and subsequent genome injection (3). In the laboratory, bacteria can readily evolve resistance to phages through mutation or loss of these surface receptors. However, in nature these surface receptors serve important functions, such as protection from antimicrobials or nutrient import, and mutation can incur fitness costs (4, 5). Given these fitness costs, bacteria in natural environments frequently rely on intracellular anti-phage defense systems, which function to restrict bacteriophage production once the phage has injected its genome into the host (6, 7).

Anti-phage defense systems have played important roles in molecular biology, as restriction-modification systems and CRISPR-Cas are powerful tools for cloning and genome editing. The discovery of CRISPR-Cas systems contributed to a renewed interest in phages and prompted investigations into the discovery of novel defense systems. A recent boon in the number of sequenced bacterial genomes coupled with new bioinformatic approaches have identified a staggering number of new anti-phage defense systems. The identification of novel defense systems has been aided in part by the observation that defense systems often cluster in bacterial genomes in so-called “defense islands” (8–10). Approaches to validate candidate defense genes typically involve cloning candidate genes into a previously phage-susceptible heterologous host and probing the engineered host with a panel of phages to assess phage sensitivity (8). Considering these newly identified defense systems, bioinformatic analyses suggest that anti-phage defense systems constitute up to 10% of a bacterium’s genome (10). Mechanistic insight into how many of these systems function to block the phage lifecycle is lacking. Of the defense systems that have been functionally characterized, many systems either target phage nucleic acids, preventing successful genome replication and packaging, and/or cause bacterial cell death, killing the host cell before phages are packaged, through what is called an “abortive infection” mechanism (7). Many of the newly discovered phage defense systems have predicted protein domains with diverse functions, implying novel mechanisms of phage defense beyond targeting nucleic acids and/or abortive infection (11).

An exciting observation stemming from the discovery of defense islands is their frequent association with mobilome genes, suggesting that bacteria can rapidly evolve phage resistance through acquisition of mobile genetic elements (MGEs) and their associated defense systems (9). However, insights into how these MGEs are mobilized and selected for in bacterial populations lags behind the discovery of these systems, as most defense systems have been studied outside of their native context on MGEs through ectopic expression in heterologous hosts (8, 11). Horizontal gene transfer is likely vital to bacterial survival in the face of rapidly evolving populations of phages, allowing previously susceptible bacteria to rapidly acquire resistance mechanisms. Transfer of phage defense systems may move with other MGE cargo genes, suggesting that selection for resistance to phages can have broad implications for bacterial fitness in response to other environmental stresses. To understand the mechanisms driving horizontal acquisition of phage defense as well as its implications for bacterial fitness requires working with naturally evolving populations of bacteria and their phages.

A promising model to address these unanswered questions in phage biology is the human pathogen *Vibrio cholerae* and its lytic phages. *V. cholerae* is an intestinal pathogen that causes substantial disease and mortality globally, particularly in regions of Southeast Asia where *V. cholerae* is endemic and causes seasonal outbreaks (12, 13). Cholera infected patients shed high titers of *V. cholerae* as well as lytic phages that propagate on their bacterial host throughout the course human disease (14). Recovering phages and *V. cholerae* from cholera patient stool samples over time can reveal the evolutionary trajectory of *V. cholerae* and its phages, and address how phage predation selects for changes in *V. cholerae*'s MGEs over time (14–16).

Surprisingly, *V. cholerae* appears to be predated upon by a low diversity of lytic phages in the context of human disease. Specifically, *V. cholerae* infected patients shed only three known lytic phages, named ICP1, ICP2 and ICP3, with ICP1 being the predominant lytic phage shed from many cholera-infected patients (14, 17). ICP1 uses the O1 antigen of *V. cholerae*'s lipopolysaccharide as its receptor to gain entry into the cell (14). In the laboratory, *V. cholerae* readily evolves mutations in the O1 biosynthesis cluster, allowing for evasion of ICP1 predation (4). However, the O1 antigen is an important virulence factor for *V. cholerae* in the context of a human host, demonstrating that escape from ICP1 predation through surface modification would result in an evolutionary dead-end for *V. cholerae* (4). However, the ongoing co-occurrence of *V. cholerae* with ICP1 is indicative of a co-evolutionary arms race, as neither entity has driven the other to extinction. This suggests that *V. cholerae* possesses diverse mechanisms that confer phage resistance other than surface modification in nature.

Genomic sequencing reveals several bioinformatically predicted phage defense systems encoded by all toxigenic (disease-causing) *V. cholerae*. Specifically, toxigenic *V. cholerae* encodes the recently described CBASS (18) and Retron (19) systems, which have been shown to inhibit non-native phages when expressed in heterologous hosts. However, all *V. cholerae* tested are successfully infected and killed by co-circulating phages, suggesting that even though predicted defense systems can be present, wild phages may encode mechanisms to overcome these systems. Alternatively, these systems may not be expressed at the appropriate time or degree to confer phage resistance. This cautions against concluding that the presence of a defense system equates with phage resistance.

To determine which resistance determinants matter in nature requires probing hosts with relevant phages. Successful phages frequently acquire mechanisms to counter defense systems, which can occur through myriad mechanisms (20). Phage-encoded counter-adaptation can occur rapidly in laboratory co-evolution experiments, which have demonstrated many routes of escape from host defenses including point mutations (21) or mimicry of host proteins (22, 23). Natural phage isolates often encode inhibitor proteins that can counter nucleic acid targeting by host-encoded restriction-modification systems or CRISPR-Cas systems; such inhibitors may function as DNA mimics (anti-restriction proteins) or as mechanistically diverse anti-CRISPR proteins (24–26). Taking the rapidness of phage counter-adaptation into account shows that to uncover defense and counter-defense in nature requires sampling of phages and their bacterial hosts across an evolutionary gradient (27). Fine scale sampling of phages and their hosts can reveal patterns of co-evolution (28). For example, when a bacterium acquires a successful anti-phage defense system, this bacterium would be able to avoid phage predation and have a higher fitness than susceptible strains. When a phage evolves a mechanism to counter this defense, this phage will then sweep through the population, supplanting sensitive phage. These sweeps are indicative of arms-race dynamics, also known as the Red Queen Hypothesis, where antagonistic entities must continually evolve to keep up with a constantly evolving antagonist (29–31).

Leveraging clinical samples of *V. cholerae* and its phages across a temporal sampling gradient revealed novel contexts for phage defense systems as discussed further in Chapter 2.

In addition to host defenses and phage-encoded counter-defenses, phage parasites add an additional layer of ecological complexity to phage-host conflicts. *V. cholerae* shed from patients frequently possesses a small genomic island which excises in response to infection by ICP1 and hijacks ICP1 replication machinery and structural genes for its own replication and packaging (32, 33). This phage parasite, PLE (phage-inducible chromosomal island like-element), lacks homology to known defense systems, yet functions to completely restrict production of ICP1 virions (32, 34). Furthermore, it does not cluster on *V. cholerae*'s genome with other known defense systems, highlighting the limitations of bioinformatic predictions of defense systems based solely on their co-occurrence with other defense genes. Despite PLE's functional similarity to defense systems (being a mobile element that inhibits phage production), PLE is fundamentally a selfish genetic element that parasitizes its viral host ICP1 for horizontal transmission. A PLE(+) infected host still lyses during infection by ICP1, yet this host only releases PLE virions, which transmit their genome horizontally in modified ICP1 virions (35). PLE virions bind to naïve neighboring cells where the PLE genome integrates into *V. cholerae*'s genome ensuring additional vertical transmission of PLE (32). Since PLE inhibits ICP1 production, it can provide population-wide protection against ICP1 infection, limiting the overall viral population and allowing for kin survival.

In this work, we address the broad question of how lytic phage predation shapes the evolution of *V. cholerae* in the context of human disease. In Chapter 2, we leverage a large collection of *V. cholerae* and phage isolates from patients across a temporal gradient to uncover patterns of co-evolution. This led to the discovery that phage defense systems are encoded on mobile genetic elements known as an SXT ICEs (integrative and conjugative elements), which are notorious for conferring antibiotic resistance in *V. cholerae* (particularly sulfamethoxazole and trimethoprim, SXT). Phage counter-adaptation to the dominant circulating SXT ICE may contribute to shifts in SXT ICE prevalence, and phage infection was found to be a stimulant for SXT ICE transfer. Chapter 3 focuses on the yet unanswered question of how PLE completely inhibits ICP1 progeny production. We uncover a nicking endonuclease with specificity for ICP1 genome sequences that inhibits ICP1 genome replication at a key stage in ICP1's lifecycle. This is the first time that a phage parasite has been shown to employ a nuclease effector to antagonize its viral host and highlights the prominent role of nucleic acid targeting in phage-host conflict. Chapter 4 explores how PLE may be protected from self-targeting by this nuclease, by encoding a putative proteolytic adaptor leading to rapid turnover of the nuclease. The work presented here contributes important insights into antagonistic co-evolution between phage and their hosts, as summarized in Chapter 5, highlighting the prominent role of mobile genetic elements in ensuring rapid adaptation to phages and phage-encoded counter-adaptation.

Chapter 2

Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts(16)

Kristen N. LeGault, Stephanie G. Hays, Angus Angermeyer, Amelia C. McKitterick, Fatema-tuz Johura, Marzia Sultana, Tahmeed Ahmed, Munirul Alam, Kimberley D. Seed

Summary

Bacteriophage predation selects for diverse antiphage systems that frequently cluster on mobilizable defense islands in bacterial genomes. However, molecular insight into the reciprocal dynamics of phage-bacterial adaptations in nature is lacking, particularly in clinical contexts where there is need to inform phage therapy efforts and to understand how phages drive pathogen evolution. Using time-shift experiments, we uncovered fluctuations in *Vibrio cholerae*'s resistance to phages in clinical samples. We mapped phage resistance determinants to SXT integrative and conjugative elements (ICEs), which notoriously also confer antibiotic resistance. We found that SXT ICEs, which are widespread in γ -proteobacteria, invariably encode phage defense systems localized to a single hotspot of genetic exchange. We identified mechanisms that allow phage to counter SXT-mediated defense in clinical samples, and document the selection of a novel phage-encoded defense inhibitor. Phage infection stimulates high-frequency SXT ICE conjugation, leading to the concurrent dissemination of phage and antibiotic resistances.

Main Text

Bacteriophage (phage) predation has been estimated to account for upwards of 40% of all bacterial mortality (36), shaping microbial community composition and selecting for bacteria that evade predation (7, 37). Co-evolution between phages and their hosts contributes to global biodiversity (7, 28, 30, 31). Indeed, genetic differences between closely related strains are largely phage resistance genes (38), with anti-phage defense systems comprising upwards of 10% of a bacterial genome (10), and mutations that decrease host susceptibility accounting for further inter-strain variability (5, 39). Phylogenetic analyses show that defense systems associate with mobilome genes, indicating that phage predation can drive bacterial evolution through mobile genetic element (MGE) flux (8, 10). A recent explosion in the identification of novel defense systems (8, 11) has allowed for a re-assessment of bacterial genomes, identifying resistance determinants in genetic contexts not previously associated with phage defense.

Despite the central role of phages in microbial evolution and ecology, there remains a dearth of biologically relevant systems to study ongoing phage-host interactions in natural environments. Approaches to study phage resistance systems often rely on ectopic expression of candidate systems in heterologous bacterial hosts and assessing restriction of phages that are not known to have encountered the system in nature. Although this approach has proven lucrative for the discovery of novel defense systems (8, 11, 40, 41) and is especially valuable for assessing genes from unculturable organisms or those without known phages, this approach precludes our understanding of fluctuating interactions and competition between defense systems, as well as the consequences and trade-offs of phage resistance. Furthermore, the extent to which a given defense system is expressed in its native context as well as the potential for dissemination of resistance to naïve hosts through horizontal transfer cannot be assessed. These approaches also limit the discovery of phage-encoded counter defenses, which lag behind our knowledge of anti-phage defenses but are likely key drivers of defense diversification.

A promising model system to address unanswered questions in phage-host ecology and evolution lies in the long-term sampling of *Vibrio cholerae* and its lytic phages (vibriophages). Pathogenic *V. cholerae* strains of the O1 serogroup possessing the key virulence factor cholera toxin are responsible for the infectious diarrheal disease cholera, which continues to pose a significant global public health threat (13). In Bangladesh, where cholera is endemic, lytic phage predation is implicated in modifying the duration and severity of cholera outbreaks (42, 43). Interestingly, cholera patients can shed both vibriophages and viable *V. cholerae*, indicating that complex interactions between phages and their hosts take place within the human intestine, as neither entity is driven to extinction. Long-term sampling of cholera-infected patients in Bangladesh has revealed the persistence of lytic phages ICP1, ICP2 and ICP3 (14), making this a powerful system to understand biologically relevant outcomes of phage predation and to discover reciprocal adaptations in phages infecting pathogenic bacteria within the context of human disease.

SXT ICEs determine clinical *Vibrio cholerae* phage susceptibility

To gain insight into the ongoing co-evolutionary dynamics between *V. cholerae* and its phages, we used time-shift assays, an established approach to reveal antagonistic co-evolution (27), to evaluate whether clinical *V. cholerae* isolates were susceptible to infection by ICP1 from past, future or contemporaneous patient samples (Figure 1.1A). A pattern emerged in which *V.*

cholerae isolates were uniformly susceptible to contemporaneous phage, but restricted ICP1 from temporally mismatched patients (Figure 1.1B). *V. cholerae* isolates from the past were susceptible to phages from this time-period, but restricted infection of phage from future patient samples, while isolates from 2018 onward restricted phages from earlier sampling dates but were susceptible to contemporaneous phages (Figure 1.1B). Importantly, *V. cholerae* DCP66 from December 2017 was uniquely susceptible to all ICP1 isolates, suggesting it lacks a phage resistance determinant, and phages from this sample were uniquely restricted by all other isolates of *V. cholerae*.

Differences in *V. cholerae*'s susceptibility to ICP1 could not be explained by the only known MGE with anti-ICP1 activity (called PLE (32)), or known anti-PLE mechanisms encoded by ICP1 (34, 44). However, comparative genomics of clinical *V. cholerae* isolates from this time-period revealed the variable presence and identity of integrative and conjugative elements (ICEs) of the SXT/R391 family (45). These ~100kb islands are of clinical importance for pathogens such as *Proteus mirabilis* and *V. cholerae*, and were originally characterized for conferring resistance to the antibiotics sulfamethoxazole, and trimethoprim (SXT), though many SXT ICEs confer an even broader suite of antibiotic resistances (46, 47). SXT ICEs are composed of core and accessory genes, with the core genes encoding functions related to SXT regulation and transmission through conjugative transfer (Figure 1.1D). Accessory genes, such as those conferring resistance to antibiotics, determine the identity of the SXT ICE and are clustered in five "hotspot" regions (46) (Figure 1.1D). SXT ICEs are named for the species of origin and country in which they were first identified, followed by a number to distinguish SXT ICEs from the same species and country (48) (e.g. *VchInd5* for "*V. cholerae* India 5"), although an identical SXT ICE can be found in multiple taxa and geographic regions (see below). The temporal pattern of phage resistance observed during this sampling period correlated with the transition of ICE*VchInd6*(+) to SXT(-) to ICE*VchInd5*(+) *V. cholerae*, suggesting that SXT ICEs may confer protection against clinically relevant vibriophages.

To test whether SXT ICEs can confer phage resistance, we conjugated ICE*VchInd6* and ICE*VchInd5* into the phage sensitive *V. cholerae* SXT(-) strain E7946. Each SXT ICE was sufficient to inhibit plaque formation by ICP1²⁰¹⁷ recovered during the sampling period that was propagated on SXT (-) *V. cholerae* (Figure 1.1C). This same ICP1²⁰¹⁷ isolate was capable of infecting the co-resident ICE*VchInd6*(+) *V. cholerae* strain when assessed for plaquing direct from stool (Figure 1.1B), but is restricted by ICE*VchInd6* when first propagated on a strain lacking an SXT ICE (Figure 1.1C), suggesting epigenetic escape (addressed further below). These results demonstrate that SXT ICEs found in *V. cholerae* natively express one or more resistance determinants sufficient to restrict vibriophages co-circulating with SXT ICEs in cholera patients.

The genes predicted to be responsible for phage resistance in ICE*VchInd5* and ICE*VchInd6* were readily identified, and in both ICEs the genes localized to hotspot 5. ICE*VchInd5* hotspot 5 contains a type 1 bacteriophage exclusion (BREX) system, an abundant phage resistance system that restricts phages through an unknown mechanism known to rely on epigenetic modification of specific sequence motifs (41, 49). Additionally, ICE*VchInd5* hotspot 5 encodes a putative abortive infection gene (Abi), and a gene of unknown function (Figure 1.1D). ICE*VchInd6* contains several restriction-modification systems as well as hypothetical genes in hotspot 5 (Figure 1.1D). Deletion of hotspot 5 from ICE*VchInd5* and ICE*VchInd6* rendered *V. cholerae* susceptible to ICP1 infection, demonstrating that the anti-phage systems in hotspot 5 are necessary for phage resistance (Figure 1.1C). On the basis of these results, we

examined hotspot 5 regions in all published SXT ICEs by supplementing data from ICEBerg 2.0 (50) with additional published SXT ICEs since the database was last updated. Strikingly, we observed that all 76 de-duplicated SXT ICEs harbor predicted anti-phage systems in the same hotspot 5 location, with the exception of atypical SXT ICEs with large rearrangements that have no identifiable hotspot 5 (51). Each hotspot 5 had at least one identifiable anti-phage system, augmenting previous *in silico* analysis of some SXT ICEs that noted the presence of cargo genes implicated in phage defense (46, 47, 52). This demonstrates that gene cargo dedicated to phage defense is a ubiquitous yet unrecognized facet of the biology of SXT ICEs (Figure 1.1D). In addition, we estimate that ~35% of the predicted ORFs found in hotspot 5 are hypothetical genes not previously linked to phage defense, suggesting a role as novel independent anti-phage systems or genes that function in concert with known defenses. All hotspot 5 regions share a putative transcriptional regulator possessing a WYL-domain that was previously identified as a core SXT ICE gene (46), yet are known to associate with defense systems (53) (Figure 1.1D). A phylogenetic tree inferred from the nucleotide sequence of the WYL-domain gene clustered SXT ICEs based on the type of anti-phage system encoded in hotspot 5, building a tree that does not agree with the relationship inferred from the conservation of the gene downstream of hotspot 5, *traI*, suggesting that hotspot 5 cargo genes move “en bloc” (Supplementary Figure 1.1). Hence, although SXT ICEs are infamous for their role in the spread of antibiotic resistance, these elements also encode for, and thus disseminate, phage resistance throughout Gammaproteobacteria, including in human pathogens like *Proteus mirabilis* and *V. cholerae* for which phage therapy/prophylaxis has been proposed (54–56).

To ascertain the global relevance of SXT ICE-mediated phage resistance in *V. cholerae*, we determined the identity and prevalence of SXT ICEs in published toxigenic O1 *V. cholerae* genomes, and in a collection of 244 isolates sequenced as a part of this study. All total, genomes of 2,600 isolates were analyzed with isolation dates going back to 1987 when SXT was first identified in sequenced O1 *V. cholerae* (12), to late 2019 (Figure 1.1E). Previous analysis of 96 SXT ICEs from seventh pandemic *V. cholerae* isolates identified a total of five different SXT ICEs in toxigenic *V. cholerae*, and found that ICEVchInd5 dominated globally (57). An additional highly similar group of SXT ICEs was identified in patient isolates from China (58), which we are unifying here as ICEVchCHN956. Our expanded analysis demonstrates that ICEVchInd5 remains globally dominant, followed by ICEVchInd6 and ICEVchBan9 (also known as ICEVchMoz10) (Figure 1.1E). Of note, a subset of strains (229/1373, ~17%) with ICEVchInd5, predominantly from Haiti, have a large deletion spanning most of hotspot 5 (Supplementary Figure 1.2). ICEVchCHN956, with genes in hotspot 5 with PFAM domains characteristic of the DISARM system (40) (Figure 1.1D), occurs in patient isolates at low abundance (Figure 1.1E). Additional SXT ICEs are found as singletons in toxigenic *V. cholerae*: MO10 (the first discovered and best characterized SXT ICE) and ICEVchInd4 (also known as ICEVchBan11) (Figure 1.1E). Despite the obvious selective advantage afforded by SXT ICEs in conferring antibiotic resistance in *V. cholerae*, we also find that SXT(-) toxigenic O1 strains persist through this sampling period.

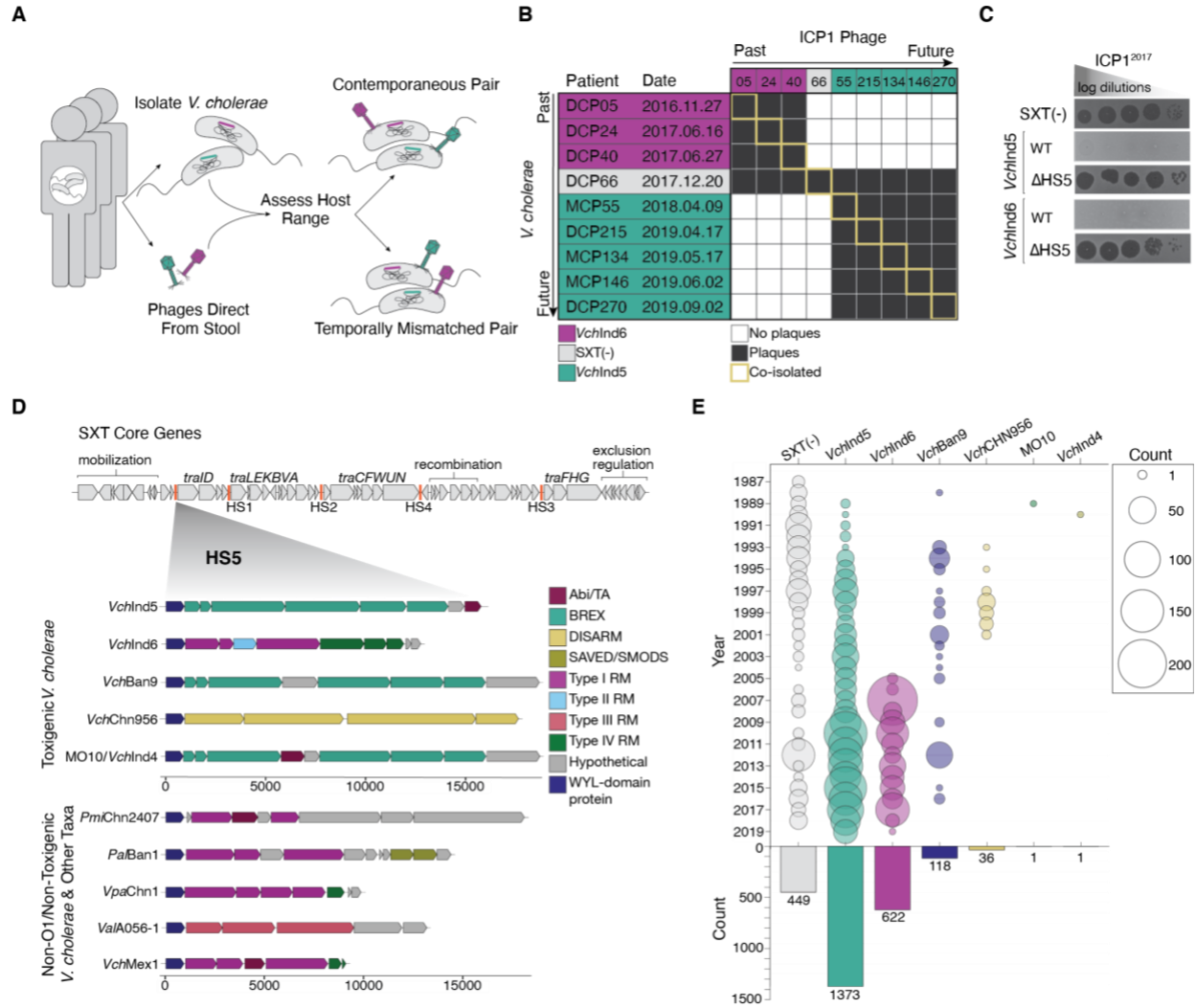


Figure 1.1. The prevalence and identity of SXT ICEs in clinical *V. cholerae* determines susceptibility to phages. (A) Schematic of time-shift experiments. ICP1 phage in cholera patient stool samples were probed for ability to productively infect isolates of *V. cholerae* from contemporaneous samples or from the past or future. Matching colors denote contemporaneously isolated *V. cholerae* and phage pairs. (B) *V. cholerae* (vertical axis) susceptibility matrix to ICP1 phages from past, present or future patients (top horizontal axis), where colors denote the presence of the SXT ICE (as identified through whole genome sequencing of patient isolates) in *V. cholerae* from the patient stool sample. ICEVchInd6 *V. cholerae* were susceptible to phages shed from patients in which ICEVchInd6 *V. cholerae* was recovered, but restricted phages from patients in which SXT(-) or ICEVchInd5 *V. cholerae* was recovered. The SXT(-) isolate was susceptible to all phages tested. ICEVchInd5 *V. cholerae* isolates restricted phages from the past but were susceptible to phages from all patients in which ICEVchInd5 *V. cholerae* was recovered. (C) Ten-fold dilutions of ICP1²⁰¹⁷ (from patient DCP24) that was propagated in the laboratory on an SXT(-) host was restricted by isogenic exconjugants of ICEVchInd6 and ICEVchInd5, and restriction was dependent on hotspot 5 (bacterial lawns in gray, zones of killing are shown in black). (D) Genomic organization of SXT ICE core genes (modified from (46)) with core function labeled above, *tra* genes encoding functions related to conjugation and hotspot regions (HS) denoted by orange lines (not to scale). Diverse anti-phage systems are encoded in hotspot 5 of SXT ICEs (shown to scale) found in toxigenic O1 *V. cholerae* (top) and other hosts (bottom). (E) Temporal distribution of SXT ICEs in toxigenic O1 *V. cholerae* between 1987-2019 (n=2,600). The size of the circle scales with the number of genomes analyzed per year, with total counts for each SXT ICE denoted below.

Genetic dissection of SXT ICE hotspot 5-mediated restriction of phages and MGEs

To gain further insight into how SXT ICEs influence *V. cholerae* ecology and gene flux, we tested the ability of SXT ICEs to restrict foreign genetic elements that are clinically important to *V. cholerae*. We focused on the three most prevalent SXT ICEs in toxigenic *V. cholerae* for further study, ICEVchInd5, ICEVchInd6 and ICEVchBan9, which contain either a BREX system, or several restriction-modification systems in hotspot 5 (Figure 1.2A). First, to assess if SXT ICEs restrict lytic vibriophages that epidemic *V. cholerae* encounters in nature, we challenged exconjugants possessing each SXT ICE with the three most prevalent lytic vibriophages found in cholera patient stool samples: ICP1 (*Myoviridae*) and unrelated podoviruses ICP2 and ICP3. Notably, a phage cocktail comprised of ICP1, ICP2 and ICP3 is under investigation to prevent cholera (54), supporting the importance of testing if SXT ICEs protect against these particular vibriophages. We included two temporally distinct isolates of ICP1, one from 2006 that is well characterized (32, 59) and one from 2017 (isolated from patient DCP24 (Figure 1.1B)). Each SXT ICE tested conferred resistance to some, but not all, of the vibriophages tested (Figure 1.2B, Supplementary Figure 1.3). Intriguingly, ICEVchInd5 restricted ICP1²⁰¹⁷ but not ICP1²⁰⁰⁶, while ICEVchBan9 restricted both phages, even though both SXT ICEs possess a BREX system. Although the mechanism of BREX restriction is not known, restriction relies on detection of unmethylated motifs (49). ICEVchInd5 and ICEVchBan9 possess divergent genes encoding for BrxX (26% amino acid identity), the essential methyltransferase and restriction determinant for BREX (41). Previously characterized BrxX proteins sharing ~40% amino acid identity were shown to modify divergent motifs (41, 49), suggesting the SXT ICE-encoded BREX systems may recognize distinct sequences. Although the recognition motif for ICEVchBan9 is not known, the putative recognition motif for ICEVchInd5, identified as RTAAAYG from previously deposited PacBio sequencing of an ICEVchInd5(+) host (accession CP033515), was abundant in both isolates of ICP1, showing that ICP1²⁰⁰⁶ has not evaded ICEVchInd5 through loss of recognition motifs.

We next wanted to determine which system is responsible for phage inhibition in *V. cholerae*, and whether different restriction enzymes restrict different phages. We deleted each restriction enzyme from ICEVchInd6, and the methyltransferases (*brxX*) from ICEVchInd5 and ICEVchBan9. ICEVchInd5 Δ *brxX* and ICEVchBan9 Δ *brxX* no longer inhibited ICP1 and ICP3, showing that the BREX system is responsible for inhibition of lytic vibriophages (Figure 1.2B). Deletion of each putative restriction enzyme in ICEVchInd6 showed that the protein with homology to BstXI (protein ID [AIM16524.1](#)), a type II restriction enzyme, is largely responsible for ICEVchInd6's anti-phage activity, however, there were differences between ICP1 isolates (Supplementary Figure 1.3A). Specifically, ICEVchInd6 Δ *bstXI* was susceptible to ICP1²⁰⁰⁶, yet ICP1²⁰¹⁷ was still partially inhibited. Deletion of the type I restriction enzyme *hsdR* also had differential effects on the two isolates of ICP1 tested. ICEVchInd6 Δ *hsdR* still completely inhibited ICP1²⁰⁰⁶, but ICP1²⁰¹⁷ was able to spontaneously overcome ICEVchInd6 Δ *hsdR* with plaquing that did not scale with the dilution series (Supplementary Figure 1.4), suggesting that HsdR has some activity against ICP1²⁰¹⁷, which notably co-circulated with ICEVchInd6(+) *V. cholerae*. ICP1²⁰¹⁷ was able to fully overcome ICEVchInd6 Δ *bstXI* Δ *hsdR*, corroborating the role for both restriction enzymes in restricting ICP1²⁰¹⁷.

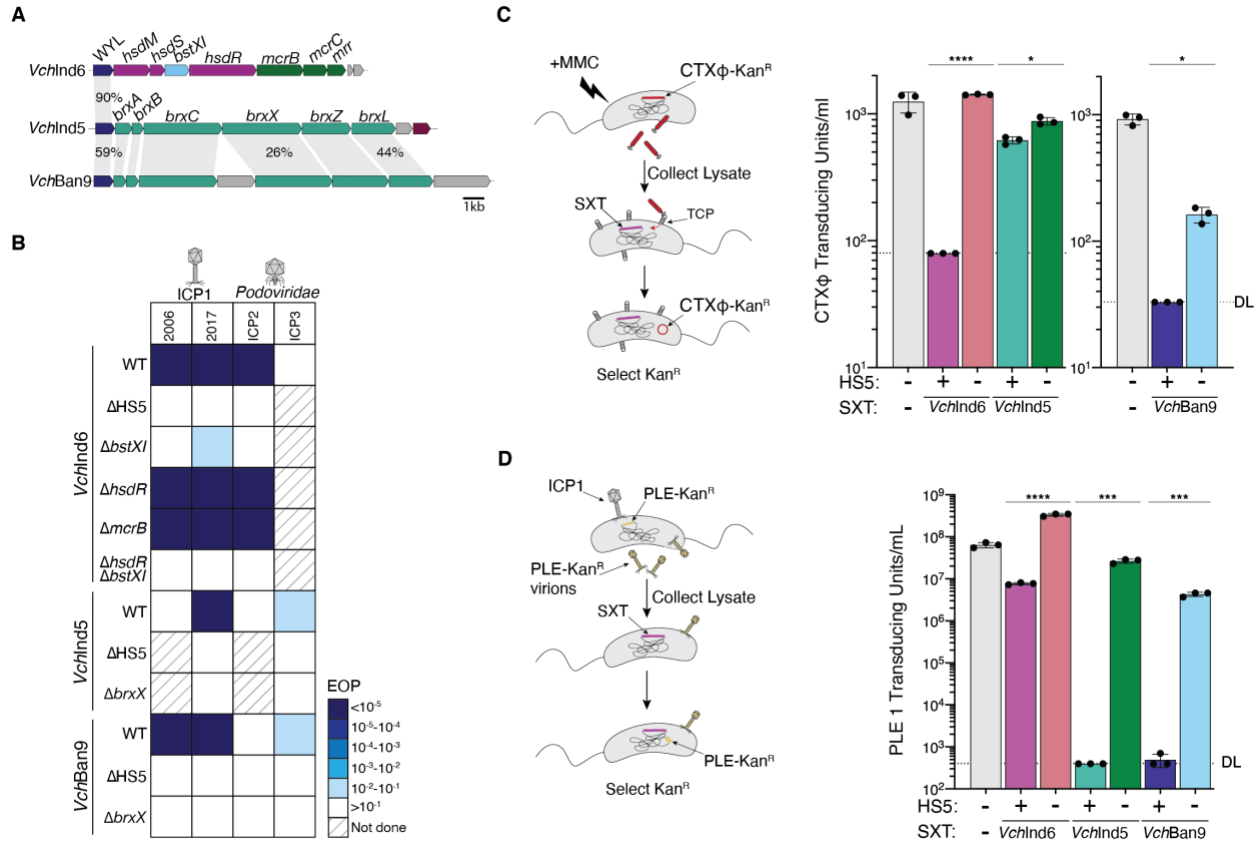


Figure 1.2. SXT ICEs confer resistance to phages and mobile elements in *V. cholerae*. (A) Genomic organization of hotspot 5 of the three dominant SXT ICEs in toxigenic O1 *V. cholerae* labeled according to homologous genes identified by REBASE/PFAM searches. Color scheme is identical to Figure 1.1D. Shading denotes proteins with the same predicted function; percent identity is indicated where the encoded proteins that share >20% amino acid identity, if unlabeled, percent identity is <20%. (B) Efficiency of plaquing (EOP) of lytic vibriophages ICP1 (isolates from 2006 and 2017), ICP2²⁰⁰⁴ and ICP3²⁰⁰⁷ on isogenic SXT ICE (+) hosts and genetic deletions, under native expression levels. When the SXT ICE did not reduce EOP, deletions were not tested. Source data is shown in Supplementary Figure 3. (C) Left: Schematic representation of the filamentous prophage CTXφ transduction assay. Right: Transduction of CTXφ into *V. cholerae* possessing each SXT ICE and derivatives with hotspot 5 deletions. Transductions into ICEVchBan9 and ICEVchBan9ΔHS5 were performed with a higher volume of CTXφ lysates than the plot to the left, resulting in a lower limit of detection (DL). Asterisks denote significance between relevant comparisons (**P* < 0.05; *****P* < 0.0001, paired t test). (D) Left: Schematic representation of PLE transduction. Right: Transduction of PLE 1 into *V. cholerae* possessing each SXT ICE and derivatives lacking hotspot 5. Asterisks denote significance of relevant comparisons (****P* < 0.001; *****P* < 0.0001, paired t test). For C and D dots represent individual biological replicates, bar height is the average; error bars display the standard deviation.

Intrigued by the conservation of WYL-domain proteins in hotspot 5 of SXT ICEs (Figure 1.1D), we assessed the necessity of these putative transcriptional regulators for SXT ICE-mediated resistance to ICP1. We observed that the WYL-domain protein was dispensable for phage resistance (Supplementary Figure 1.5A-C). However, ectopic expression of each WYL-protein in their cognate SXT ICE(+) background greatly restored the sensitivity of SXT ICE(+) *V. cholerae* to ICP1 (Supplementary Figure 1.5D-F). These data are consistent with the WYL-domain proteins acting as repressors for hotspot 5, though the conditions under which the WYL-domain proteins are important for regulating anti-phage defenses remains to be investigated.

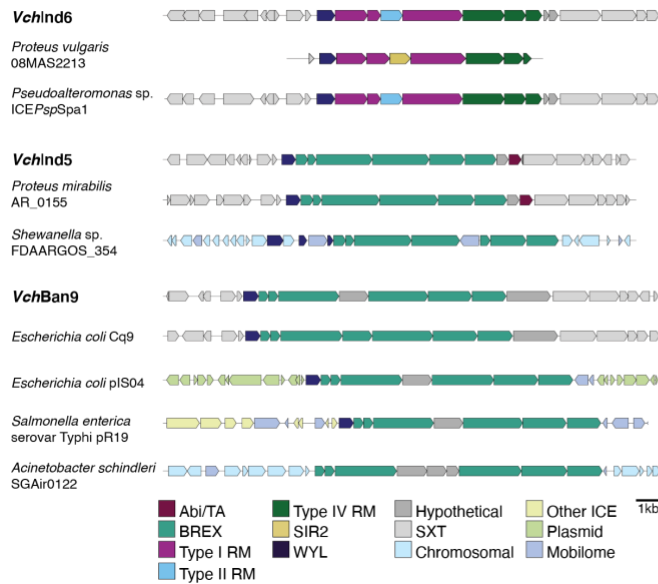
Following on the surprising discovery that SXT(-) *V. cholerae* persist in clinical samples (Figure 1.1B and 1.1E), we postulated that SXT ICEs may incur an fitness cost to their host by restricting the acquisition of beneficial MGEs. To this end, we assessed SXT-ICE-mediated restriction of two clinically important and beneficial MGEs in epidemic *V. cholerae*: the temperate phage CTX ϕ and the phage-inducible chromosomal island like elements (PLEs). The filamentous phage CTX ϕ encodes cholera toxin, and is essential for *V. cholerae*'s pathogenicity in the human host (60). PLEs are ~20kb phage satellites that confer resistance to ICP1 and have been a persistent part of the mobilome of epidemic *V. cholerae* dating back to at least 1949, before SXT was common in clinical *V. cholerae* isolates (32). We assessed the acquisition of antibiotic marked CTX ϕ and each PLE by otherwise isogenic *V. cholerae* hosts either lacking an SXT ICE or harboring each of the most prevalent SXT ICEs (ICEVchInd6, ICEVchInd5, ICEVchBan9), or derivatives lacking hotspot 5 (Figure 1.2C and D). SXT ICEs restricted acquisition of both PLEs and CTX ϕ , but to varying extents, and in all cases, transduction was at least partially restored in mutants lacking hotspot 5 (Figure 1.2C and D and Supplementary Figure 1.6). PLE 1 was uniquely restricted by ICEVchInd6 compared to PLEs 2, 3, 4 and 5 (Figure 1.2D and Supplementary Figure 1.6). To explain this discrepancy, we looked for the presence of the characterized motif for BstXI (CCANNNNN[^]NTGG) from *Bacillus subtilis* (61) (which shares ~50% amino acid homology with ICEVchInd6's BstXI) and found the motif once in PLE 1 but not in the other PLEs. To determine if this motif is indeed cleaved by ICEVchInd6's BstXI, crude lysates from an *Escherichia coli* host ectopically expressing *bstXI* were assessed for DNA cleavage activity using a PCR product spanning the predicted cleavage site from PLE 1 as a substrate. Lysates containing BstXI, but not from the empty vector control, cleaved the PCR product, generating two cleaved fragments of the anticipated size (Supplementary Figure 1.7A). A PCR product with the predicted motif mutated was not cleaved by BstXI lysates (Supplementary Figure 1.7B). Corroborating these findings, knocking out the gene in the PLE 1 that contains the BstXI recognition motif restored PLE 1 transduction into ICEVchInd6 to the level observed for the SXT(-) host (Supplementary Figure 1.7C). Collectively, these data demonstrate that like other anti-phage defense systems, SXT-mediated restriction of MGEs comes at a cost. While all SXT ICEs restricted infection by clinically relevant lytic vibriophages, they also limited acquisition of CTX ϕ and PLEs, which provide beneficial traits to epidemic *V. cholerae*. These observations help to explain the maintenance of SXT(-) strains in nature (Figure 1.1E), and the loss of hotspot 5 in some SXT ICE(+) isolates (Supplementary Figure 1.2).

SXT ICEs disseminate phage resistance to other Gammaproteobacteria

SXT ICEs are recombinogenic, self-transmissible mobile elements hypothesized to evolve by genetic exchange of variable and hotspot regions (57, 62) suggesting that gene cargo in hotspot 5 could be mobilized between taxa through both SXT ICE conjugation and through recombination of hotspot 5 with other SXT ICEs or homologous DNA. To determine the extent of lateral dissemination, genes in hotspot 5 from ICEVchInd6, ICEVchInd5 and ICEVchBan9 were queried using BLASTn to probe whether the genes could be found in other taxa, either within SXT ICEs or in a different genetic context. The genes in ICEVchInd6 hotspot 5 were always localized to an SXT ICE, and ICEVchInd6 in its entirety is found in other Gammaproteobacteria with over 99% nucleotide identity (Figure 1.3A, Supplementary Table 1.2). Intriguingly, a variant of ICEVchInd6 was also identified in *Pseudoalteromonas* sp. referred

to as ICEPspSpa1 (63). ICEPspSpa1 is 96.45% identical to ICEVchInd6 except in place of *bstXI* is a SIR2 domain protein implicated in the Thoeris defense system (8) (Figure 1.3A). Despite being globally dominant in toxigenic *V. cholerae* O1 strains, the ICEVchInd5 BREX system was rarely found outside of *V. cholerae*, occurring in hotspot 5 from an otherwise divergent SXT ICE in *P. mirabilis* and outside of an MGE in *Shewanella* (Figure 1.3A and Supplementary Table 1.3). The BREX system from ICEVchBan9 was the most widely distributed, both taxonomically and in terms of genetic context, being found within SXT ICEs (for example, in *E. coli* and *P. mirabilis*), as well as outside of SXT ICEs. These varied locations include the chromosome, on plasmids, as well as in non-SXT ICEs (Figure 1.3A and Supplementary Table 1.4). The distribution of hotspot 5 encoded anti-phage systems in other Gammaproteobacteria suggests that these systems confer resistance to a broad range of phages infecting different hosts well beyond *V. cholerae*, and that the self-transmissible nature of SXT ICEs disseminates these anti-phage systems to new taxa.

A



B

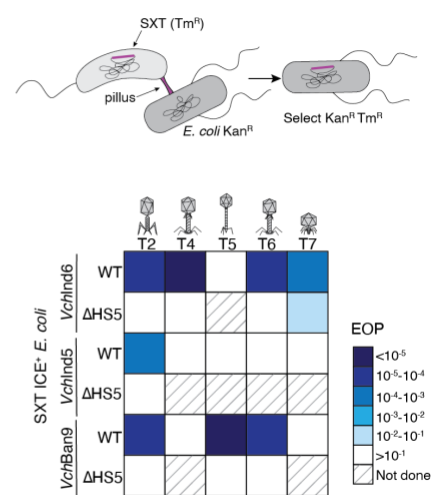


Figure 1.3. Hotspot 5 from *V. cholerae* SXT ICEs occur in other Gammaproteobacteria and confer phage resistance upon conjugation into *E. coli*. (A) Gene maps (shown to scale) depicting representative instances of hotspot 5 contents from *V. cholerae* SXT ICEs occurring in other taxa (labeled to the left) and contexts, indicated by the color-coded flanking genes and labeled in the shaded box to the right. Hotspot 5 from ICEVchInd6, ICEVchInd5 and ICEVchBan9 along with flanking SXT ICE genes are shown for reference. The full list of hotspot 5 contents found in other taxa is included in Supplementary Tables 1.2-1.4. Genes were identified using BLASTn (shown are hotspot 5 genes with >90% nucleotide identity across >75% of the query). (B) Top: Schematic depicting SXT ICE conjugative transfer from *V. cholerae* into *Escherichia coli*. Below: Efficiency of plaquing (EOP) of lytic coliphages on isogenic SXT ICE(+) *E. coli* and corresponding hotspot 5 deletions under native expression levels. When the SXT ICE did not reduce EOP, the deletion was not tested. Source data is shown in Supplementary Figure 8.

To test whether SXT ICEs can confer phage resistance upon entry into a new host, ICEVchInd6, ICEVchInd5 and ICEVchBan9 and the cognate Δ hotspot 5 derivatives, were conjugated into *E. coli* (Figure 1.3B) and exconjugants were challenged with the lytic coliphages T2, T4, T5, T6 and T7. Strikingly, all SXT ICEs tested offered protection from at least one coliphage, with the broadest protection afforded by ICEVchInd6 (restricting all phage except for T5) and the narrowest restriction afforded by ICEVchInd5 (which only partially restricted T2) (Figure 1.3B,

Supplementary Figure 1.8). ICEVchBan9, which is natively found in *E. coli* (Figure 1.3A), restricted T2, T5 and T6, while T4 and T7 were not inhibited. The T7-encoded DNA mimic protein, Ocr, was recently shown to inhibit type 1 BREX (25) and thus may explain T7's resistance to ICEVchBan9 and ICEVchInd5. *E. coli* exconjugants harboring the Δ hotspot 5 SXT ICE derivatives were susceptible to all phages tested, demonstrating that it is the anti-phage systems in hotspot 5 that confer resistance in *E. coli* and *V. cholerae*. In line with this, experimental investigation of an SXT ICE (ICEVspPor3) from *Vibrio splendidus* encoding a type III restriction modification system in hotspot 5 conferred resistance to coliphage T1 upon conjugation into *E. coli* (64). Collectively, these results demonstrate that upon acquisition of SXT ICEs through conjugation, hotspot 5 is natively expressed in the recipient taxa and confers resistance to their cognate phages.

Mechanisms in vibriophage ICP1 to counter SXT-ICE-mediated phage restriction

In light of the discovery that SXT(+) *V. cholerae* restrict phages through anti-phage systems known to rely on differential epigenetic modification and subsequent restriction, we postulated that phages shed from patients infected with SXT ICE(+) *V. cholerae* have escaped restriction through stochastic acquisition of DNA modification. Epigenetically modified phage can then productively infect co-circulating strains possessing the same SXT ICE but are unable to infect *V. cholerae* with a different SXT ICE (as shown in our time-shift assay, Figure 1.1B). In further support of this, phage in the patient stool in which we recovered SXT ICE(-) *V. cholerae* DCP66 were restricted by all SXT ICE(+) hosts, and the SXT ICE(-) *V. cholerae* isolate from this patient was uniquely susceptible to infection by phages from all patient stool samples (Figure 1.1B). Other genetic means of overcoming modification-based restriction are common among phages and include anti-restriction proteins and loss of recognition motifs (reviewed in (20)). To distinguish between potential mechanisms of epigenetic or genetic escape, ICP1 isolates from patient stool samples were first passaged on SXT ICE(-) *V. cholerae*, and then evaluated for the ability to productively infect the clinical *V. cholerae* strains from our original time-shift experiment (Figure 1.4A). ICP1 isolates from stool samples in which we co-recovered ICEVchInd6(+) *V. cholerae* lost the ability to productively infect isolates with ICEVchInd6, suggesting that ICP1 shed in patient stool samples had escaped through acquiring modification during the course of infection (Figure 1.4A). In contrast, the pattern of susceptibility seen with ICEVchInd5 and cognate phages indicated a genetic route of escape. Specifically, ICP1 shed from patient stool samples with *V. cholerae* harboring ICEVchInd5 retained the ability to infect co-circulating ICEVchInd5(+) isolates after passage through an SXT(-) host, yet ICP1 isolates shed in patient samples in which we recovered ICEVchInd6(+) *V. cholerae* were still restricted by ICEVchInd5 (Figure 1.4A). This was reminiscent of the differential phenotypes observed for the two ICP1 isolates tested on ICEVchInd5: ICP1²⁰⁰⁶ was not restricted by ICEVchInd5, whereas ICP1²⁰¹⁷ was restricted (Figure 1.2B).

To identify the genetic mechanism responsible for the host range differences between ICP1 isolates, we compared the genomes of ICP1 isolates from across a temporal gradient that can and cannot infect *V. cholerae* harboring ICEVchInd5 (Figure 1.4B). This analysis revealed differences in the region of the genome encoding gene products (gp) 21-25 that co-varied with plaque formation on a ICEVchInd5(+) host (Figure 1.4B and 1.4C). Unlike ICP1 isolates that can productively infect ICEVchInd5(+) *V. cholerae*, the restricted ICP1²⁰¹⁷ lacks both *gp21* and the promoter region for *gp25*, which we identified by analyzing the ICP1²⁰⁰⁶ transcriptome (59)

(Figure 1.4C and Supplementary Figure 1.9A), suggesting one of these early genes may allow ICP1 to overcome restriction by ICEVchInd5. Having narrowed in on these two gene products as candidate anti-ICEVchInd5 proteins, we evaluated if *in trans* expression of either *gp21* or *gp25* was sufficient to rescue plaque formation for the restricted ICP1²⁰¹⁷ phage. Ectopic expression of *gp25*, but not *gp21*, restored plaque formation by ICP1²⁰¹⁷ on *V. cholerae* harboring ICEVchInd5 (Figure 1.4D), supporting our designation of *gp25* as *orbA* for “overcome restriction by BREX”. To decipher if OrbA was the sole factor permitting ICP1 plaque formation on the ICEVchInd5(+) host, we expressed *orbA* during ICP3 infection, since ICP3 is an unrelated phage that is also restricted by the BREX system in ICEVchInd5 (Figure 1.2B). Here too we observed that *orbA* expression restored productive infection for ICP3 on the ICEVchInd5(+) host (Figure 1.4D). Additionally, ICP1²⁰⁰⁶ $\Delta orbA$ was no longer able to productively infect ICEVchInd5(+) *V. cholerae* (Supplementary Figure 1.9C), validating that *orbA* is both necessary and sufficient to overcome ICEVchInd5’s BREX system. These results demonstrate that OrbA (YP_004250966) functions as an anti-BREX protein with activity against ICEVchInd5. OrbA encodes a small 44 amino acid protein with no homology to any annotated proteins, as well as no PFAM/DUF domains. Furthermore, PSI-BLAST did not return any significant homologs outside of ICP1. The T7-encoded anti-restriction protein Ocr has been shown to overcome BREX in *E. coli* by functioning as DNA mimic (25). Although the crystal structure for OrbA has not been solved, a predicted model does not show the electrostatics characteristic of a DNA mimic like Ocr (Supplementary Figure 1.10), suggesting a different mechanism of action for OrbA’s anti-BREX activity. Notably, ICEVchInd5 does not restrict ICP2, CTX ϕ nor the majority of *E. coli* phages tested, even though these genomes possess the putative ICEVchInd5 methylation site RTAAAYG. Although the mechanism of BREX restriction is not known, restriction relies on detection of unmethylated motifs (49). This suggests the other unrestricted phages and MGEs tested here either fail to activate BREX or encode mechanisms to overcome ICEVchInd5, though they do not have homologs of *orbA*. Our findings also indicate that OrbA is not a broadly active anti-BREX protein since the BREX system encoded by ICEVchBan9 equally restricts ICP1²⁰⁰⁶ and ICP1²⁰¹⁷, isolates that express or do not express *orbA*, respectively (Figure 1.2B). The finding that the BREX system in ICEVchBan9 is insensitive to OrbA activity suggests that diversification of BREX systems in SXT ICEs may be driven by the need to avoid phage-encoded inhibitors.

We next set out to document the reciprocal dynamics between the repertoire of SXT ICEs in *V. cholerae* with anti-BREX activity afforded by OrbA in ICP1 in greater detail. Towards this end, we assessed the presence and identity of SXT ICEs in 523 isolates of toxigenic O1 *V. cholerae* from Bangladesh spanning 2001-2019. During this time-period we observed fluctuating dominance of ICEVchInd5 and ICEVchInd6 and a low level of SXT ICE(-) *V. cholerae* (Supplementary Figure 1.11). To probe the relationship between phage predation and transitions in SXT ICE prevalence, we analyzed a subset of 148 *V. cholerae* and 44 ICP1 isolates from November 2016 to September 2019 collected and sequenced from cholera patient samples as a part of this study (Figure 1.4E). As was evident in the subset of samples in which ICP1 and *V. cholerae* were co-isolated (Figure 1.1B), we observed a transition between circulating ICEVchInd6(+) strains to ICEVchInd5(+) strains in 2018, with intermittent recovery of SXT ICE(-) *V. cholerae* from patients between June 2017 and May 2018. Of note, SXT(-) *V. cholerae* did not possess additional resistance determinants, as all SXT(-) isolates from the sampling period were broadly sensitive to the phages tested, with the exception of some isolates possessing a receptor mutation which has been shown previously to render *V. cholerae*

insensitive to ICP2 (5) (Supplementary Figure 1.12). Interestingly, ICP1 isolates lacking the *orbA* promoter region (as in ICP1²⁰¹⁷ (Figure 1.4C)) were only isolated during a time-period when ICEVchInd5 was absent in cholera patients (Figure 1.4E); however, concurrent with the emergence of *V. cholerae* ICEVchInd5(+) in patient stool samples in the spring of 2018, all contemporary ICP1 isolates possess the promoter for *orbA*. This locus provides an example of selection for a novel anti-BREX gene found in phage, which can counter restriction by an SXT ICE, and shows how temporal dynamics in SXT ICEs can exert purifying selection on phages. In times when ICP1 is not encountering ICEVchInd5 in nature, the locus that confers resistance to ICEVchInd5 is not under positive selection. Thus, variants with promoter mutations persist, or are perhaps selected for, suggesting this locus provides no fitness benefit to ICP1 when it is not co-circulating with ICEVchInd5(+) *V. cholerae*.

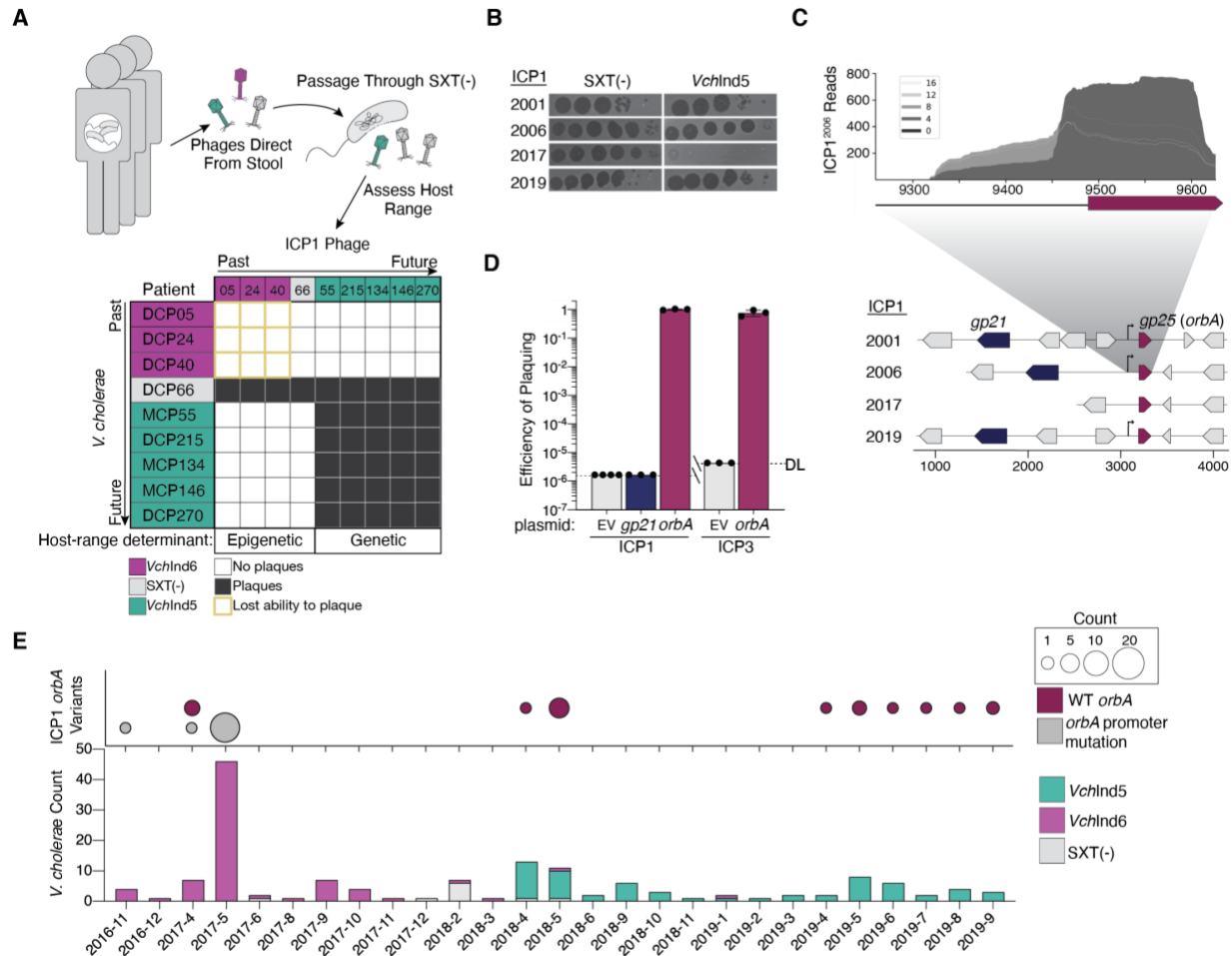


Figure 1.4. Clinical ICP1 overcomes co-circulating SXT ICEs through acquisition of epigenetic modification as well as an anti-BREX protein OrbA. (A) ICP1 isolates from cholera-patient stool samples were propagated on an SXT(-) strain and then probed for ability to infect clinical *V. cholerae* isolates. ICP1 isolates from ICEVchInd6(+) stool samples lost the ability to infect co-circulating *V. cholerae* upon passage through an SXT(-) host indicative of initial escape through epigenetic modification. ICP1 isolates from ICEVchInd5(+) patients retained infectivity of co-circulating *V. cholerae* after passage through an SXT(-) host, suggestive of a genetic means of overcoming ICEVchInd5. (B) Tenfold dilutions of laboratory-passaged ICP1 isolates spotted on SXT(-) *V. cholerae*, or an isogenic exconjugant of ICEVchInd5, identified ICP1²⁰¹⁷ (from patient DCP24) as being uniquely inhibited by ICEVchInd5 (bacterial lawns in gray, zones of killing are shown in black). (C) Comparative genomics of ICP1 isolates showing that ICP1²⁰¹⁷ lacks *gp21-24* and the promoter for *gp25*, as identified through RNA-seq of ICP1²⁰⁰⁶. Top: Average read coverage across the indicated genome position covering the predicted promoter and

gp25, reads are color coded by time point in minutes post infection. (D) Efficiency of plaquing (EOP) of ICP1²⁰¹⁷ and ICP3 in the presence of ectopically expressed *gp21*, *gp25* (*orbA*) or an empty vector control (EV) in ICEVchInd5(+) *V. cholerae*: *gp25* (designated *orbA* for “overcome restriction by BREX”) but not *gp21* restores plaquing. A greater inhibition of ICP3 was observed compared to Figure 1.2B, possibly due to the growth defect incurred by maintaining antibiotic selection obscuring the very small plaques that ICP3 forms in the presence of ICEVchInd5 (see Supplementary Figure 9B). Dots display independent biological replicates and bar height indicates the mean with error bars showing standard deviation. The dashed line indicates the limit of detection (DL). (E) Sampling and whole genome sequencing of 148 *V. cholerae* isolates and 44 ICP1 isolates from cholera patient stool samples in Bangladesh, depicting isolate counts (y-axis) over time, showing the transition from ICEVchInd6 to SXT(-) to ICEVchInd5(+) *V. cholerae*. ICP1 isolates with the *gp25* promoter mutation (details shown in Supplementary Figure 1.9A) were only isolated from patients in which we recovered ICEVchInd6(+) *V. cholerae*, and have not been isolated since the re-emergence of ICEVchInd5(+) *V. cholerae*.

Phage infection stimulates SXT ICE conjugation

SXT ICEs are SOS responsive, meaning that DNA damage leading to the accumulation of single stranded DNA triggers de-repression of SXT, culminating in SXT excision and expression of operons related to conjugative transfer (65). DNA damage may serve as a proxy for impaired cell viability, thus SXT ICEs that can mobilize and spread horizontally to escape a compromised host would have a selective advantage. DNA damaging agents such as mitomycin C (MMC), ultraviolet radiation, and certain classes of antibiotics that interfere with DNA replication (for example, ciprofloxacin), have been shown to increase SXT conjugation frequency (66). Despite these studies, ciprofloxacin is still taken by cholera patients in endemic regions (67). Because phages including ICP1 are being investigated to prevent cholera (54), we sought to assess whether phage infection would similarly promote SXT ICE conjugation and thus inadvertently promote the spread of antibiotic resistance. To this end, we tested the frequency of conjugative transfer of ICEVchInd5 into recipient *V. cholerae* following infection by either ICP1²⁰⁰⁶ or ICP1²⁰¹⁷, representing phages that are not restricted or are fully restricted by ICEVchInd5, respectively (Figure 1.2B). ICEVchInd5(+) *V. cholerae* was infected with ICP1 at a multiplicity of infection of 0.5, and frequency of transfer into a differentially marked *V. cholerae* recipient was determined. Importantly, the added differentially marked *V. cholerae* recipient lacks the O1 antigen, which is the receptor for ICP1 (14), making the recipient insensitive to ICP1 infection, yet proficient for conjugation (Supplementary Figure 1.13). As expected, conjugation frequency was elevated above background following treatment with MMC (Figure 1.5A). Interestingly, infection by ICP1²⁰⁰⁶ resulted in high frequency SXT ICE transfer compared to an uninfected control, resulting in a conjugation frequency comparable to what was observed from a MMC treated donor (Figure 1.5A). In contrast, neither the restricted phage ICP1²⁰¹⁷ nor addition of cell lysate from uninfected *V. cholerae* (Supplementary Figure 1.14C), enhanced conjugation. Transfer frequency was remarkably constant despite productive phage infection by ICP1²⁰⁰⁶ among the donor strains, resulting in a 100-fold decrease in donor cell viability (Supplementary Figure 1.14A-B). As expected, no antibiotic resistant exconjugants were detected with a $\Delta traD$ donor, demonstrating that phage-mediated transduction is not a contributing factor in these assays (Figure 1.5A). To build on this observation, we tested whether infection by coliphage could also stimulate higher frequency conjugation of ICEVchInd5 from an *E. coli* donor to a *V. cholerae* recipient. Consistent with the results for *V. cholerae*, we observed that infection by an unrestricted phage, T7, also resulted in high frequency transfer (Figure 1.5B). It is unclear whether it is the infected cells that are conjugating at a greater frequency, or if the uninfected cells in the population are sensing phage lysis through an unknown mechanism

and becoming high frequency donors. We propose alternative models, the first where cellular damage wrought by phage infection triggers increased transfer of SXT ICEs as the SXT ICE races to transfer out of a dying host (Figure 1.5C), or, alternatively, productive phage infection in neighboring cells leads to increased conjugative transfer from uninfected cells, through a yet undescribed mechanism. Regardless of mechanism, these data highlight that phages promote the dissemination of SXT ICEs and associated antibiotic and phage resistances.

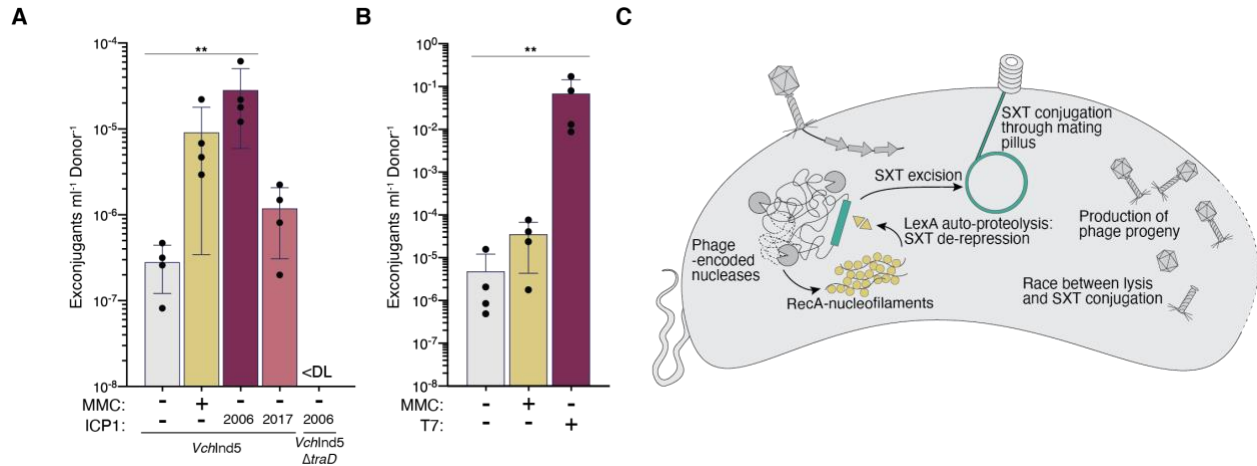


Figure 1.5. ICP1 infection increases SXT ICE conjugation frequency. (A) Conjugation frequency from a donor *V. cholerae* mock treated (LB), mitomycin C (MMC) treated or infected by ICP1²⁰⁰⁶ (unrestricted phage) or ICP1²⁰¹⁷ (restricted phage), into a phage-insensitive ΔO1 *V. cholerae* recipient. No exconjugants were detected from a phage infected donor lacking *traD* (DL is the limit of detection). Asterisks denote significance (**P<0.01, one-way ANOVA, Tukey HSD, where no asterisks are displayed no significant differences were detected). (B) Conjugation frequency from donor *E. coli* into *V. cholerae*, where the donor was mock treated (LB), mitomycin C (MMC) treated or infected by the lytic phage T7 (which is unrestricted by ICE *VchInd5*). Asterisks denote significance (**P<0.01, Kruskal-Wallis, Dunn's test was performed as variances were unequal, where no asterisks are displayed no significant differences were detected). For A and B, dots represent individual biological replicates, bar height is the average; error bars display the standard deviation. (C) Model representing a possible mechanism of ICP1-stimulated SXT conjugation: Upon infection, ICP1-encoded nucleases degrade the *V. cholerae* genome, triggering SXT de-repression through the SOS response, resulting in the SXT ICE escaping the lysing host cell through conjugation.

Discussion

ICEs of the SXT/R391 family appeared in sequenced epidemic *V. cholerae* in 1987 (12) and are now appreciated as a prominent genomic feature of seventh pandemic strains (49, 50). SXT ICEs are infamous for conferring multidrug resistance (45, 46, 57), however, diverse SXT ICE cargo genes can confer a broad suite of adaptive traits to their bacterial hosts (46, 70–73). Here we show that SXT ICEs invariably encode phage-resistance genes, linking together antibiotic and phage resistances on a single mobile genetic element widespread in Gammaproteobacteria. Phage defense systems cluster in hotspot 5, while antibiotic resistances are encoded by variable region III and hotspot 3 (46), suggesting that hotspot regions may have dedicated functions, and that other accessory loci could be probed for conserved functionality. The mosaicism of SXT ICEs is well appreciated (46) and is known to be driven by recombination between SXT ICEs via a RecA-independent SXT-ICE-encoded recombination system (62). Genetic exchange between hotspots and bacterial chromosomes, plasmids and other ICEs may be frequent, as we find near nucleotide identical hotspot 5 anti-phage systems in

contexts outside of SXT ICEs. The potential for rampant genetic exchange suggests SXT ICEs can rapidly adapt to keep up with their rapidly evolving phage antagonists.

Co-evolution between *V. cholerae* and lytic vibriophages is driven, at least in part, by presence and identity of SXT ICEs. Using a time-shift experiment, we find that lytic phage can infect contemporary *V. cholerae* hosts but cannot infect hosts from the past or the future. Observed at this short time scale, co-evolution of ICP1 with *V. cholerae* appears to be driven by selective sweeps. Adaptations that allow for phage infection of contemporary SXT ICE(+) *V. cholerae* sweep to fixation over short times scales, rendering contemporary ICP1 more fit than ICP1 from the near past, characteristic of arms race dynamics (74, 75). ICP1's adaptations to overcome contemporary SXT ICEs is driven both by changes in its capacity to express the anti-ICEVchInd5 BREX inhibitor OrbA and acquisition of epigenetic modification allowing for infection of ICEVchInd6(+) hosts. The heterogeneity in ICE SXT identity and presence in circulating toxigenic *V. cholerae* suggests the potential for population bottlenecks and subsequent sweeps. Phages that have acquired anti-SXT ICE mechanisms will sweep through the population, providing a selective advantage to hosts with rare SXT ICEs that confer resistance to the dominant phages. The subsequent sweep of the resistant SXT ICE through the *V. cholerae* population generates a further bottleneck on the ICP1 population. These dynamics of fluctuating genomic dominance in a population of bacteria and phages allows for apparent long-term coexistence between phages and their bacterial hosts, where neither entity drives the other to extinction. The heterogeneity in SXT ICE-mediated phage defense can be considered a “pan-immune system” (7), where the collective pool of SXT ICEs circulating in *V. cholerae* serve to restrict all of the dominant lytic phages: ICP1, ICP2 and ICP3. Additionally, these temporal dynamics provide insight into the ongoing maintenance of SXT(-) *V. cholerae*, which should have a selective disadvantage within the clinical context since they lack the antibiotic resistances conferred by SXT ICEs. However, SXT(-) *V. cholerae* provide a population-wide benefit, as phage that have propagated on an SXT(-) host would lack appropriate epigenetic modifications for subsequent infection of SXT ICE(+) hosts, allowing for persistence of *V. cholerae*. The prevalence of diverse anti-phage systems carried by other SXT ICEs suggests that these dynamics play out in other Gammaproteobacteria.

BREX is the dominant anti-phage system encoded by SXT ICEs in toxigenic O1 *V. cholerae*, being encoded by both ICEVchInd5 and ICEVchBan9. Despite the observation that BREX is found in ~10% of bacterial genomes (41), BREX function had not previously been probed for anti-phage activity in its native context. Nonetheless, previous studies uncovered important aspects of how BREX functions, showing that the methyltransferase, BrxX, is required for hemi-methylation of host DNA at non-palindromic motifs and that non-methylated DNA is restricted (41, 49). However, the mechanism of BREX restriction is not known, though phage DNA is thought not to be degraded but somehow unable to replicate (49). Coliphage that escape restriction by *E. coli*'s BREX system can be isolated during passaging in the laboratory and were shown to have acquired methylation (49). We were unable to generate ICP1 epigenetic escapes in the laboratory on either ICEVchInd5(+) nor ICEVchBan9(+) hosts, suggesting *V. cholerae*'s BREX systems are more restrictive to their cognate phages. Epigenetic escapes were also not observed in patient stool samples in which ICEVchInd5(+) *V. cholerae* was recovered. Our study corroborates previous findings that BREX restriction varies widely for different phages. This may be explained by the number of recognition motifs in a phage genome, though as we show here, anti-BREX inhibitors play additional roles in BREX's inferred anti-phage activity. Probing

SXT ICE(+) *V. cholerae* harboring natively-expressed BREX systems with lytic vibriophages may be a valuable model for elucidating lingering questions about BREX function. The ability of SXT ICEs to restrict beneficial mobile genetic elements such as CTX ϕ and PLEs, coupled with the previous finding of an SXT ICE-encoded endonuclease that limits natural transformation (70), has implications for bacterial evolution via horizontal gene transfer. Each SXT ICE tested restricted transduction of some, but not all, elements tested. The prevalence of a restrictive SXT ICE in a population, which may fluctuate in response to phage pressure, could therefore impact the relative abundance of other MGEs, further demonstrating the extent to which phage predation can alter the pool of genes in a population. The ability for SXT ICEs to restrict MGEs and phages shows a tradeoff that may favor SXT(-) *V. cholerae* in certain conditions, such as in changing environments (76) where acquisition of novel traits may be favored. SXT ICEs create a barrier to genetic exchange from cells possessing a different SXT ICE (or no SXT ICE), as only DNA from a donor with the same modification-based restriction system will be seen as ‘self’ and thus unrestricted. Indeed, comparative genomics has shown that genetic exchange between genomes is higher if they encode cognate RM systems (77).

The human gut may represent an environment where SXT ICE conjugation can occur between gut commensals or invading pathogens, especially when DNA damaging antibiotics such as ciprofloxacin have been administered, as ciprofloxacin has been shown to increase SXT conjugation through activation of the SOS response (66) and is detected in cholera patients (67). The appearance of SXT ICEs in toxigenic *V. cholerae* in the late 1980s rendered previously used drugs such as tetracycline and trimethoprim/sulfamethoxazole largely ineffective for treating cholera patients (78). Continued antibiotic use may select for the emergence of new resistances in *V. cholerae*, perhaps through SXT ICE exchange between taxa. An even greater suite of drug resistances, such as ciprofloxacin resistance, has been attributed to SXT ICEs found in *Proteus mirabilis* (47), which possesses near identical SXT ICEs to those found in toxigenic *V. cholerae*. Concerns over antibiotic use has prompted some to suggest lytic phages as a means to treat or prevent cholera (54, 79). However, we have shown here that phage infection can also increase the frequency of SXT conjugation from *V. cholerae* and *E. coli*. Interestingly, infection by a phage that is not restricted by ICEVchInd5 stimulated higher frequency conjugation, while infection by a restricted phage had no discernable effect on conjugation. There is an expected energetic cost and accompanying growth delay associated with the formation of the conjugative machinery, as has been shown for other ICEs (80). When infected by a restricted phage, there may be little benefit for SXT ICEs to trigger conjugation, as their host cell may suffer a selective disadvantage limiting the SXT ICE’s vertical transmission. However, during productive phage infection, an SXT ICE that can escape a dying cell would be able to move laterally. As we continue to grapple with increasing multidrug resistant bacterial infections, which have garnered a renewed interest in phage therapy, it is of critical importance that we work to understand how phages drive the evolution of dangerous pathogens like *V. cholerae*. When specifically considering the application of phages for infections caused by Gammaproteobacteria, in which SXT ICEs can be abundant, caution should be taken as phages can simultaneously spread the genes for antibiotic and phage resistances by stimulating the transfer of SXT ICEs. Selection of phages for therapeutic use and their preparation should also take into account SXT ICE-mediated phage restriction. Continued surveillance of epidemic *V. cholerae* and predatory vibriophages will promote a molecular understanding of phage-bacterial coevolution in the context of human disease, which is necessary to advance efforts to manipulate phage-bacterial interactions for therapeutic or prophylactic benefit

Materials and Methods

Strains and growth conditions

Stool samples were collected following institutional review with written consent from participants or for guardians of young participants. Stool samples collected from suspected cholera patients attending icddr, Dhaka hospital, and Government Health Complex of Mathbaria, Pirojpur were screened for *V. cholerae* serogroup O1 and/or O139 using Crystal VC Rapid Dipstick test (RDT; Span Diagnostics, Surat, India). RDT-positive stool samples were de-identified, given a number, mixed with glycerol (to 20% final volume), stored at -80°C and shipped to University of California, Berkeley for further analysis. For isolation of *V. cholerae* onsite in Bangladesh, the RDT-positive stool samples were enriched in alkaline peptone water (APW) (pH-8.4, Difco, Detroit, MI) for 6-8 h at 37°C; and *V. cholerae* was isolated by culturing overnight on taurocholate tellurite gelatin agar (TTGA) (Difco). Typical *V. cholerae*-like colonies on TTGA were selected and confirmed using a combination of biochemical and serological methods, as described previously (81).

Additional isolation of *V. cholerae* and phages from stool was performed at the University of California, Berkeley. Initial isolation of *V. cholerae* was completed by streaking from stool directly on thiosulfate-citrate-bile salts-sucrose agar and Vibrio ChromoSelect agar (Sigma). Stool samples were also used to inoculate liquid outgrowth in APW at 30°C and 37°C with samples struck out from liquid cultures on thiosulfate-citrate-bile salts-sucrose agar and Vibrio ChromoSelect agar plates 6 and 24 hours after inoculation. Colonies appearing to be *V. cholerae* were colony purified at least twice on agar plates before confirmation by PCR and subsequent whole genome sequencing. These *V. cholerae* isolates were then used to isolate phages. For phage isolation, *V. cholerae* hosts were grown to mid-log, exposed to stool or APW outgrowths, and plated on 0.5% LB top agar. Plaques were picked and plaque purified twice before confirmation by PCR and/or whole genome sequencing.

A detailed list of strains used in this study is found in Data S1. Isogenic strains of each SXT ICE were generated by conjugations into *V. cholerae* E7946 or *Escherichia coli* MG1655 backgrounds (described below). 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: 32 µg/ml trimethoprim, 75 µg/ml kanamycin, 100 µg/ml spectinomycin, 100 µg/ml streptomycin, 2.5 µg/ml chloramphenicol (*V. cholerae* on plates), 1.25 µg/ml chloramphenicol (*V. cholerae* in liquid), 25 µg/ml chloramphenicol (*E. coli*). Plasmid constructs were induced with 1mM isopropyl-D-1-thiogalactopyranoside (IPTG) and 1.5mM theophylline (Sigma) for 20 minutes prior to plaque assays (described below), and inducers were maintained in the agar overlays. For BstXI production from *E. coli*, strains were induced as above for 4 hours at 37°C.

Phage spot plates were performed as a visual aid but not used to quantify phage titers, and were performed as described previously (82). The efficiency of plaquing (EOP) for each phage was calculated by comparing the number of plaques that a given phage forms on the SXT ICE (-) host relative to the number of plaques formed on an otherwise isogenic SXT (+) host. Each EOP was calculated as the average of assays done in biological triplicate, and the limit of detection is the point at which the phage is unable to productively infect the SXT (+) host (<20 plaques at the highest concentration tested) while still forming plaques on a SXT (-) host. For these assays bacteria, *V. cholerae* or *E. coli*, were grown to mid-log, phage were added and given 10 minutes

for adsorption at room temperature before the mixture was added to molten 0.5% LB agar (with the exception that T7 plaques were enumerated using 0.8% LB agar) with each dilution plated on a single petri dish.

Bacterial mutants were generated through natural transformation of linear DNA created by splicing by overlap extension PCR (83). Antibiotic resistance cassettes were resolved from deletions following expression of the flp-recombinase leaving an in frame frt scar. Plasmids were constructed with Gibson Assembly using a modified pMMB67EH derivative vector (82). Phage mutants were selected as previously described (84). All mutations and constructs were verified by Sanger sequencing.

Determining phage-host range with phages directly from stool

A small amount of frozen stool sample collected on a pipette tip was added to 50 μ l of APW, which was then serially diluted and plated with *V. cholerae* isolates from patient stool samples using phage spot assays (82). The presence/absence of plaquing in Figure 1B was determined from two biological replicates of this assay with all hosts. To verify the identity of the phages from patient stool samples, plaques were picked from both biological replicates and resuspended in 50 μ l of STE. 10 μ l of the phage-STE sample was boiled for 10 minutes and used as template to verify the identity of phages by PCR using unique primer sets (Supplementary Table 1.1) to differentiate ICP1, ICP2 and ICP3. Only samples with ICP1 were used for the time-shift assay. For Figure 4A, ICP1 plaques from stool were propagated on *V. cholerae* E7946 (SXT(-)). Three plaques were picked from each original passaged plaque and these phages were probed for the ability to form plaques by spot assay on the panel of *V. cholerae* isolates from patient stool samples.

Conjugations and transductions

Transfer of SXT ICEs from was performed as described previously (66). Briefly, overnight donor cultures were grown in LB supplemented with trimethoprim and recipient strains were grown with appropriate antibiotics (either kanamycin or spectinomycin). Equal volumes (500 μ l) of donor and recipient cultures were added to a 1.5mL microcentrifuge tube and pelleted at 5000 $\times g$ for 3 minutes. Pellets were resuspended in 50 μ l LB and pipetted onto either an LB agar plate or a 0.8 μ m filter disk placed on top of an LB agar plate. Conjugations were allowed to proceed for 6 hours (hr) at 37°C, after which time the plate was flooded with 1mL of LB and a dilution series was plated on either trimethoprim and kanamycin or trimethoprim and spectinomycin to select for recipients that acquired the SXT ICE. Mock conjugations were performed which excluded either donor or recipient to ensure that no spontaneous double antibiotic resistant colonies appeared. The presence of the entire SXT ICE in the *V. cholerae* E7946 background was further verified by whole genome sequencing of at least two exconjugants for each ICEVchInd6, ICEVchInd5 and ICEVchBan9.

Phage or mitomycin C (MMC) stimulated conjugations were performed with overnight cultures of donor ICEVchInd5(+) *V. cholerae* (either KL571/572) or *E. coli* (KL684) split into four 1mL tubes and treated as follows: mock infected (added appropriate volume of phage diluent), phage at a multiplicity of infection (MOI)~0.5, or MMC (20ng/mL). Following addition of treatments, all tubes were incubated at 37°C with aeration for 15 minutes. After 15 minutes, 0.5mL of the treated culture was mixed with 0.5mL of a differentially antibiotic marked recipient lacking the

O1 antigen (KL322) and conjugations proceeded as described above. The filter was placed into 1mL of LB and the tube was vortexed for 30 seconds to detach cells. An aliquot of cells was plated on selective agar to quantify total number of viable donor cells at the end of the conjugation. To select for and enumerate exconjugants, cells were plated on trimethoprim with the appropriate antibiotic and enumerated the next day.

CTX transductions were performed as described previously (59) with the following modifications: CTX ϕ production from the donor strain was induced for 4 hours and Δ ctxAB recipient strains engineered to overexpress the CTX receptor (TCP) were induced for 4 hours. Recipient strains were either SXT(-), SXT ICE(+) as well as their Δ hotspot 5 derivatives. CTX lysates were added to recipient strains at a ratio of 1:4. For ICEVchInd6 and ICEVchInd5 and their derivative recipients, 250 μ L of lysate was added to 750 μ L of recipient. Initial experiments indicated that CTX ϕ was still restricted by ICEVchBan9 Δ HS5, so to test for a more subtle difference, volumes were scaled up to 600 μ L of lysate and 1800 μ L of recipient. Transductants were plated on LB plates supplemented with kanamycin and spectinomycin to select for CTX(+) recipients and enumerated following overnight incubation.

PLE transductions were performed as described previously (35). Briefly, an SXT(-) donor strain with a PLE possessing a kanamycin cassette was grown to OD₆₀₀=0.27-0.32 in LB supplemented with 10mM MgCl₂ and infected with ICP1²⁰⁰⁶ at an MOI=2.5. Cultures were washed twice to remove unbound phage and resuspended in fresh LB with 10mM MgCl₂. Following lysis, cultures were treated with chloroform to remove un-lysed cells. Lysate was combined at a 1:1 ratio with overnight cultures of an SXT(-) recipient (KS882), or each SXT ICE(+) as well as their Δ hotspot 5 derivatives and incubated at 37°C for 1 hour. Transductants were plated on LB plates supplemented with kanamycin and spectinomycin to select for PLE (+) recipients and enumerated following overnight incubation.

Preparation of crude lysates and DNA cleavage assays

Crude lysates were prepared following a modified protocol from Gordeeva et al. 2019 (49). 25mL cultures of *E. coli* expressing either *bstXI* or an empty vector control were induced for 4 hours. Following induction culture flasks were placed on ice and all subsequent steps were performed on ice or at 4°C. Cultures were pelleted at 4000 x *g* and the pellet was re-suspended in 1mL of lysis buffer: 150mM Tris-Cl pH=7.8, 50mM NaCl, 1 Pierce Protease Inhibitor mini tablet (ThermoFisher Scientific), 7mM BME and 1mM EDTA. Resuspended pellets were sonicated for 5 minutes total “on” time, in pulses of 10 seconds “on” and 20 seconds “off”. Lysed cell debris was removed by centrifugation at 10,000 x *g* for 3 minutes. To remove excessive contamination of DNA, the lysates were passed through a DNA binding column (Syd Labs), by centrifugation for 1 minute at 10,000 x *g*. The flow-through was collected and kept on ice. To perform cleavage assays, PCR products were obtained by amplifying the region within PLE 1 containing the putative BstXI recognition motif, or a plasmid harboring the mutant derivative (CCAATATTATGG to ACAATATTACGG). 100ng of each DNA product was incubated in separate reactions with increasing amounts of lysates (2 μ L, 5 μ L or 10 μ L) and the volume was adjusted to 20 μ L final by addition of reaction buffer: 150mM Tris-Cl pH=7.8, 50mM NaCl, 10mM MgCl₂, 0.1mg/ml bovine serum albumin (BSA). Reactions were carried out in PCR tubes for 30 minutes at 37°C, then heat inactivated at 80°C for 5 minutes before running the entire reaction on a 0.8% agarose gel and visualized by staining with GelRed.

Whole genome sequencing and Genomic Analysis

Bacterial DNA was prepared for whole genome sequencing using commercially available kits (Qiagen® DNeasy Blood and Tissue Kit and/or Monarch® Genomic DNA Purification Kit (New England BioLabs)). High titer phage stocks were prepared as previously described: through either polyethylene glycol precipitation or concentration on centrifugal filters (15), before DNase treatment to remove contaminating *V. cholerae* DNA and genomic DNA extraction of phage DNA using commercially available kits (Qiagen® DNeasy Blood and Tissue Kit and/or Monarch® Genomic DNA Purification Kit from New England BioLabs). The manufacturer's protocol was modified by doubling the reagent volumes to account for the volume of phage stock used as starting material. DNA libraries were prepared with commercial kits (NEBNext Ultra II or NEBNext Ultra II FS DNA Library Preparation) or performed by the Microbial Genome Sequencing Center (Pittsburgh, PA). Sequencing (150-by-150 paired end) was performed by the QB3 Genomics Core at the University of California, Berkeley, or by the Microbial Genome Sequencing Center. Accession numbers for *V. cholerae* and ICP1 genomes sequenced as part of this study are included in supplemental files Data S3 and S4.

Computational approaches

To determine the prevalence of hotspot 5 genes from ICEVchInd5, ICEVchInd6 and ICEVchBan9 in taxa other than *V. cholerae*, the nucleotide sequence spanning the start of the WYL-domain to the end of the final ORF before the start of *traI* (Figure 1.1D) was queried against the NCBI non-redundant nucleotide database using the BLASTn algorithm in October 2020. All hits in non-*V. cholerae* taxa with greater than 75% query coverage and 90% nucleotide identity were considered homologous.

Phylogenetic trees were constructed from nucleotide alignments using MUSCLE of the conserved WYL-domain containing gene and the conjugative relaxase *traI*. Trees were built using PhyML with 100 bootstrap iterations, and visualized using FigTree v1.4.4.

A database of *V. cholerae* genomes was constructed by downloading all assemblies from the relevant NCBI genome directory (<https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/505/>) and by a literature search for sequenced *V. cholerae* genomes (Data S3). These accessions were subsequently downloaded from the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) in fastq format and assembled with spades v3.14.0 using default settings (85). The database was refined to include only toxigenic O1 *V. cholerae* as determined by being positive for both *ctxB* (accession: CP001235 region: 1646195-1646569) and O1-antigen (accession:LT906614; region: 245143-278060) by BLASTn with $\geq 80\%$ length and $\geq 80\%$ identity. Representative SXT ICEs (Data S5) and their extracted hotspot 5 contents were then queried against every contig with ≥ 1 kb length and $\geq 5\times$ coverage (if coverage was known) from every genome assembly using BLASTn. The total length of each BLASTn hit was summed and the percent identities were averaged for each SXT. The SXT variant with the highest length was considered to be present in a genome if its total hit length was $\geq 80\%$ of the SXT's sequence length and the average percent identity was $\geq 90\%$ of the SXT sequence. Because many SXT ICEs that were previously identified as unique would not be separated by our cutoffs, we included an additional analysis where we took our curated list of both full-length and hotspot 5s from Data S5 and performed a BLASTn search

against 51 additional SXT ICEs and have summarized which SXT ICEs would not have been distinguished by our analysis in Data S6.

Accession numbers and hotspot 5 boundaries are included in supplemental Data S2. All complete SXT ICE sequences were downloaded from ICEBerg 2.0 (50) and the hotspot 5 region was identified by looking for the conserved flanking genes *s025* and *traI*. Additional SXT ICEs not included in the ICEBerg 2.0 database were identified by looking for homologs of the TraI protein sequence from ICEVchInd6 (QAY18233.1), ICEVchInd5 (ACV96327.1), ICEVchBan9 (ACQ61870.1), MO10 (AAL59675.1) and ICEPmiUSA1 (CAR44884.1) using PSI-BLAST with default settings in May and June 2020. Significant matches to the TraI protein sequence were then assessed for the presence of the conserved SXT ICE genomic organization (46) and manually curated to avoid duplication of known SXT ICEs. The type of anti-phage system contained in hotspot 5 was determined by searching for conserved domains using PFAM (86) search as well as looking for methyltransferase and restriction enzyme homologs using REBASE (87). Gene maps for Figure 1.1D, 2A, 3A and S1A were made using the package gggenes in R version 4.0.0. MO10 and ICEVchInd4 share >96% nucleotide identity across hotspot 5, so a single gene map was shown for the two SXT ICEs. Statistical analyses were performed using GraphPad Prism 9.0.2 Software, LLC. A predicted model of OrbA was generated using the program i-TASSER (88).

Acknowledgments:

The authors are especially thankful to icddr,b hospital and lab staff for support. The authors thank members of the Seed lab for critical feedback and thoughtful discussion regarding this manuscript and Vivek Mutalik for sharing the *E. coli* phages.

Chapter 3

A phage parasite deploys a nicking nuclease effector to inhibit viral host replication(89)

Kristen N. LeGault, Zachary K. Barth, Peter DePaola IV, Kimberley D. Seed

Summary

PLEs are phage parasites integrated into the chromosome of epidemic *Vibrio cholerae*. In response to infection by its viral host ICP1, PLE excises, replicates and hijacks ICP1 structural components for transduction. Through an unknown mechanism PLE prevents ICP1 from transitioning to rolling circle replication (RCR), a prerequisite for efficient packaging of the viral genome. Here, we characterize a PLE-encoded nuclease, NixI, that blocks phage development likely by nicking ICP1's genome as it transitions to RCR. NixI-dependent cleavage sites appear in ICP1's genome during infection of PLE(+) *V. cholerae*. Purified NixI demonstrates *in vitro* nuclease activity specifically for sites in ICP1's genome and we identify a motif that is necessary for NixI-mediated cleavage. Importantly, NixI is sufficient to limit ICP1 genome replication and eliminate progeny production, representing the most inhibitory PLE-encoded mechanism revealed to date. We identify distant NixI homologs in an expanded family of putative phage parasites in *Vibrios* that lack nucleotide homology to PLEs but nonetheless share genomic synteny with PLEs. More generally, our results reveal a previously unknown mechanism deployed by phage parasites to limit packaging of their viral hosts' genome and highlight the prominent role of nuclease effectors as weapons in the arms race between antagonizing genomes.

Introduction

Biological conflicts represent a major selective pressure shaping microbial diversity from the level of genes to communities (28, 38). Nowhere are these conflicts more antagonistic than between bacteria and the lytic viruses (phages) that infect them. Bacterial genomes contain abundant anti-phage defense systems that function through diverse mechanisms to restrict phages. In addition to the selective pressure imposed by an evolving repertoire of host immunity, phages themselves are hosts to their own subcellular parasites. Phage parasites, often referred to as phage satellites, may appear to be defunct prophages, as they are integrated into specific attachment sites in their bacterial hosts' genomes (90), yet phage parasites conspicuously lack the full suite of structural genes required for virion assembly (91). In response to proteins encoded by their viral hosts, phage parasites excise (92), replicate (93) and steal structural components from their viral hosts to package their own genome, and upon lysis of the cell, disseminate their genome to naïve neighboring cells in modified virions (94, 95). To accomplish this, phage parasites encode mechanisms to inhibit packaging of their viral host's genome and favor packaging of the parasite genome. These mechanisms must balance inhibition of their viral host's lifecycle with the phage parasites' requirement for products encoded by the host virus. In this way, phage parasites employ finely tuned inhibitory mechanisms to redirect viral host resources for the parasites own use. Much of our understanding of the mechanisms that phage parasites use to subvert their viral hosts comes from the well-characterized phage-inducible chromosomal islands (PICIs) (91, 96), of which, *Staphylococcus aureus* pathogenicity islands (SaPIs) are the archetypical members (97). However, diverse families of unrelated phage parasites have likely evolved independently (98) and though often overlooked, such parasites are likely ubiquitous within bacterial genomes (99). Understanding diverse phage satellites can help define mechanistic paradigms of phage parasitism and uncover novel innovations not seen in model systems interrogated thus far.

PLEs (phage-inducible chromosomal island-like elements) are phage parasites unrelated to SaPIs that have proven lucrative for uncovering new mechanisms of parasitism (15, 35, 100). Based on current data, PLEs are restricted to toxigenic *Vibrio cholerae*, the causative agent of the diarrheal disease cholera. PLEs appear to exclusively parasitize the lytic phage ICP1, which is also frequently shed in cholera-patient stool samples (14). PLEs have been identified in *V. cholerae* genomes dating back to the 1940s, and to date 10 distinct PLEs have been identified that share considerable nucleotide identity (32, 33). All PLEs are ~17kB in size and encode ~26 open reading frames (ORFs) that largely lack predicted functions or known PFAM domains. Surveillance efforts indicate that PLE1 is the predominant PLE in contemporary *V. cholerae* isolates from cholera patients in Bangladesh, where cholera is endemic (15, 32). For this reason, PLE1 has been the subject of mechanistic studies aimed at understanding PLE activity (35, 82, 100, 101) and will generally be referred to here as 'PLE' unless another variant is specified. Although PLE is known to completely abolish ICP1 production (32, 34), the mechanisms PLE uses to inhibit ICP1 remain somewhat elusive. PLE employs multiple strategies to inhibit ICP1 progeny production, as no single ORF knockout allows ICP1 to overcome PLE (100). To antagonize PLE in nature, ICP1 isolates encode anti-PLE effectors: either a CRISPR-Cas system that targets the PLE genome (34) or a site-specific nuclease directed to the PLE origin of replication (102). Such phage-encoded anti-PLE nucleases speak to the fitness costs ICP1 faces when PLE activity is left unimpeded.

Like other known phage parasites such as SaPIs, PLE modifies its viral host's capsid, constraining ICP1's capsid to ~1/3 of its normal size to accommodate the smaller PLE genome and exclude ICP1's larger ~125 kB genome (35). Beyond redirecting virion morphogenesis proteins to package its own genome and restrict ICP1, recent work shows that PLE uses a multi-pronged approach to attack and compete with ICP1 DNA replication, both by reducing total ICP1 genome copy and by inhibiting a key step in ICP1's genome replication program (101). PLE replication requires the PLE-encoded replication initiation factor RepA and hijacking of ICP1's replication machinery to drive a nearly ~1000-fold increase in genome copy during ICP1 infection (32, 101). Concurrent with robust levels of PLE replication, PLE reduces ICP1's DNA replication ~four-fold, suggesting a link between PLE replication and inhibition of ICP1 replication (101). In addition to restricting ICP1's genome replication, PLE also alters the mode of ICP1 replication (101). At 8 minutes post-infection (mpi) of a permissive PLE(-) host, ICP1 replication proceeds through bidirectional theta replication (101). Later in infection, between 12 and 16 mpi, ICP1 transitions to rolling circle replication (RCR) (101). RCR is a common mode of replication for viral genomes and plasmids (103–105). Continued rounds of RCR generates linear genome concatemers, an essential step for many viruses because linear concatemers are the DNA substrate recognized by the terminase complex that initiates DNA packaging into capsids (106). When ICP1 infects a PLE(+) *V. cholerae* host, early bidirectional theta replication proceeds normally, however, ICP1's transition to RCR is inhibited by an unknown mechanism (101). Importantly, even when PLE is unable to replicate (as in PLE (+)Δ*repA*) and total ICP1 replication is partially restored, ICP1's transition to RCR is still inhibited, suggesting that an additional PLE-encoded factor impedes ICP1's transition to RCR (101). To our knowledge, PLE is unique among phage parasites in targeting genome replication of its viral host, highlighting this stage of the viral lifecycle as a ripe arena for uncovering new mechanisms of interference. Here, we work to uncover the PLE-encoded mechanism that inhibits ICP1's transition to RCR. We identify aberrant cut sites in ICP1's genome that appear during infection of a PLE(+) host and identify the PLE-encoded nuclease necessary for generating these cut sites. This PLE-encoded nuclease, referred to as NixI, is expressed coincident with ICP1's attempt to transition to RCR in a PLE(+) host. We show that NixI inhibits ICP1 genome replication, nearly abolishing virion production. We identify a sequence motif that is necessary for NixI-mediated cleavage and provide evidence that PLE is not only protected from NixI-mediated cleavage, but actually requires NixI-mediated cleavage of ICP1 to achieve wildtype levels of PLE replication. We further identify homologs of the PLE-encoded nuclease in what we consider to be an expanded family of putative phage parasites in non-toxigenic *V. cholerae* and other non-cholera *Vibrio* species.

Results

ICP1's genome is cleaved during infection of PLE(+) *V. cholerae*

While assessing ICP1 replication profiles by mapping Illumina read coverage for each nucleotide position (referred to from here on out as DNA-seq) (101), we noted regions in ICP1's genome where reads coverage dropped precipitously during infection of PLE (+) but not PLE(-) *V. cholerae*. These coverage drops suggested potential sites of nuclease cleavage. In order to characterize these putative cut sites more systematically we drew on a previously published method that identified cleaved sequences in transcripts from RNA-seq data in order to assess

activity of an endoribonuclease toxin (107). To generate cleavage maps across ICP1's genome, we calculated the coverage ratio of ICP1 reads 16 minutes post-infection of PLE(+) relative to PLE(-) *V. cholerae*. This time point is late in the infection cycle, at a time when ICP1 and PLE are both replicating (101) and before the onset of lysis ~20 minutes post-infection of PLE(+) *V. cholerae* (32, 100) (Figure 2.1A).

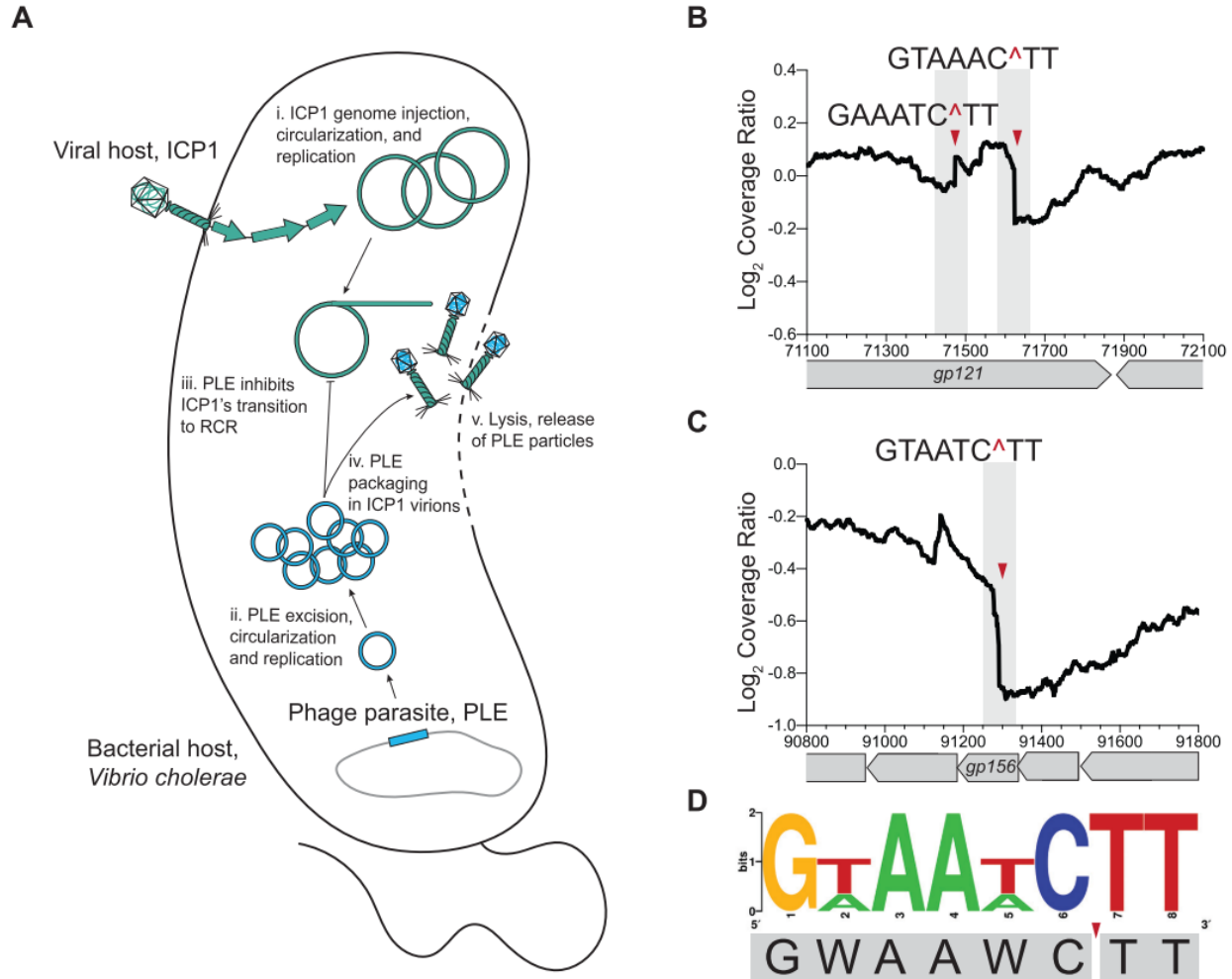


Figure 2.1. ICP1's genome exhibits signs of cleavage when infecting PLE(+) *V. cholerae* (A) Model of ICP1 infection a PLE(+) *V. cholerae*: The viral host ICP1 injects its genome, triggering PLE excision and gene expression. Both ICP1 and PLE replicate their genomes initially, yet midway through infection PLE prevents a further increase in ICP1 genome copy and inhibits ICP1's transition to rolling circle replication (RCR). PLE hijacks ICP1 structural proteins for packaging, and no ICP1 progeny are produced. *V. cholerae* lyses ~20 minutes post-infection releasing only PLE transducing particles. (B) Cleavage ratio PLE(+)/PLE(-) across ICP1's genome position (x-axis) showing the most prominent putative cut site within ICP1's *gp121*, with the two putative cut sites indicated by the red triangles and the nucleotides flanking the cut region above. The cleavage ratio is calculated as the average coverage of three biological replicates. (C) As in B, but displaying the other most prominent cut region within ICP1's *gp156*. Additional cut sites are summarized in Supplemental Supplementary Figure 1A. (D) WebLogo displaying the consensus motif for cut sites in *gp121* and *gp156* with the cut site indicated by the red triangle.

We identified a total of 63 candidate cut sites in ICP1 whose appearance varied $\pm$ PLE with $\log_2$ coverage drops $>(0.026)$, with several coverage drops clustering in the same region of ICP1's genome (Materials and Methods, Supplemental Supplementary Figure 2.1A). We

identified coverage drops across ICP1's putative packaging (*pac*) site that were present during infection of a PLE(-) host but were reduced in a PLE(+) host (Supplementary Figure 2.1A and 2.1B). The *pac* site is recognized and cleaved by the terminase complex to initiate genome packaging, and a lack of cutting in a PLE(+) host is consistent with PLE-mediated inhibition of ICP1 genome packaging. The remaining putative cut sites were coverage drops that appeared in ICP1's genome when infecting a PLE(+) host but not a PLE(-) host, suggesting ICP1's genome is aberrantly cut when infecting a PLE(+) host. Two sites in ICP1's genome, located in *gp156* and *gp121*, showed the most striking drops in coverage, while additional sites showed more subtle drops in coverage including a second site in *gp121* located about 60bp 5' of the dominant coverage drop (Figure 2.1B and 2.1C, Supplementary Figure 2.1C). We therefore chose to focus on the sites in *gp156* and *gp121* in greater detail. We identified a shared motif based on the two sites in *gp121* and single site in *gp156* of GWAAWC[^]TT (W=A or T), where the "[^]" denotes the position of potential cleavage (Figure 2.1D). Where multiple coverage drops occurred within a 100bp region (as occurred for many regions, Supplemental Supplementary Figure 1A), the sequence at the most dominant coverage drop was used as input to generate the consensus motif for the PLE-dependent putative cut sites for a total of 17 sites. These sites generated a less stringent motif GNAANC[^]TT that occurs 99 times in ICP1, of which 16 sites displayed a coverage drop at the 16-minute time point (Supplementary Figure 2.1D). The sites in ICP1's genome that have the motif but were not cleaved may be protected due to protein occupancy or were perhaps missed because this assay was performed at just one time point. The analysis of reads coverage across ICP1's genome suggests a limited number of pronounced coverage drops that appear only when infecting a PLE(+) host, consistent with nucleolytic cleavage at these regions.

ICP1's genome is targeted by a PLE-encoded nicking endonuclease, NixI

We hypothesized that the putative cleavage sites in ICP1 apparent during infection of PLE(+) *V. cholerae* were caused by a PLE-encoded nuclease. However, we lacked a candidate nuclease, as the majority of PLE-encoded open reading frames (ORFs) have no primary sequence similarity to any annotated genes. Using the more sensitive HHpred, which searches for remote homology based on predicted structural domains, we identified a single PLE-encoded ORF, NixI (WP_032468646), that shares similarity with the intron-encoded homing endonuclease I-HmuI (probability score=97.1%, e-value=0.0058) (108). Despite sharing only 10% amino acid identity across the protein, NixI shares an HNH/N catalytic motif with I-HmuI (Figure 2.2A). No other proteins shared significant homology with NixI by HHpred, which only compares proteins with solved structures. I-HmuI is the only crystallized protein from a family of closely related nicking HNH/N endonucleases (108, 109), and a MUSCLE alignment including these proteins identified several conserved residues at the catalytic core, including an experimentally validated catalytic asparagine from I-HmuI at position 95 in NixI (108) (Figure 2.2A). To address whether *nixI* may encode for a nuclease, we cloned *nixI* or a mutated *nixI*^{N95A} allele, which we postulated would be catalytically inactive, under an inducible promoter in *V. cholerae* and plated for cell viability with and without inducer. Consistent with *nixI* functioning as a nuclease, induction completely abolished *V. cholerae* colony formation, as the predicted recognition motif we identified in ICP1's genome, (GWAAWC[^]TT, Figure 2.1D), is found 638 times in *V. cholerae*'s chromosomes. In contrast, induction of NixI^{N95A} had no impact on colony formation (Figure 2.2B), indicating that toxicity of NixI is associated with nucleolytic activity.

Having identified a putative PLE-encoded nuclease, we next asked whether *nixI* is required for the appearance of the putative cleavage sites in ICP1's genome during infection of PLE(+) *V. cholerae*. We infected PLE $\Delta nixI$ *V. cholerae* with ICP1 and again performed DNA-seq to evaluate changes in cleavage ratios of PLE(+) $\Delta nixI$ relative to *V. cholerae* PLE(-). Strikingly, we observed that all coverage drops in ICP1's genome seen during a PLE(+) infection were abolished during infection of a PLE(+) $\Delta nixI$ host (Figure 2.2C and Supplementary Figure 2.1C), indicating that *nixI* is necessary for cleavage of ICP1's genome *in vivo*. However, cleavage at ICP1's *pac* site was still inhibited in a PLE(+) $\Delta nixI$ host, indicating a PLE-encoded mechanism of inhibiting ICP1's DNA packaging independent of *nixI* (Supplementary Figure 2.1B).

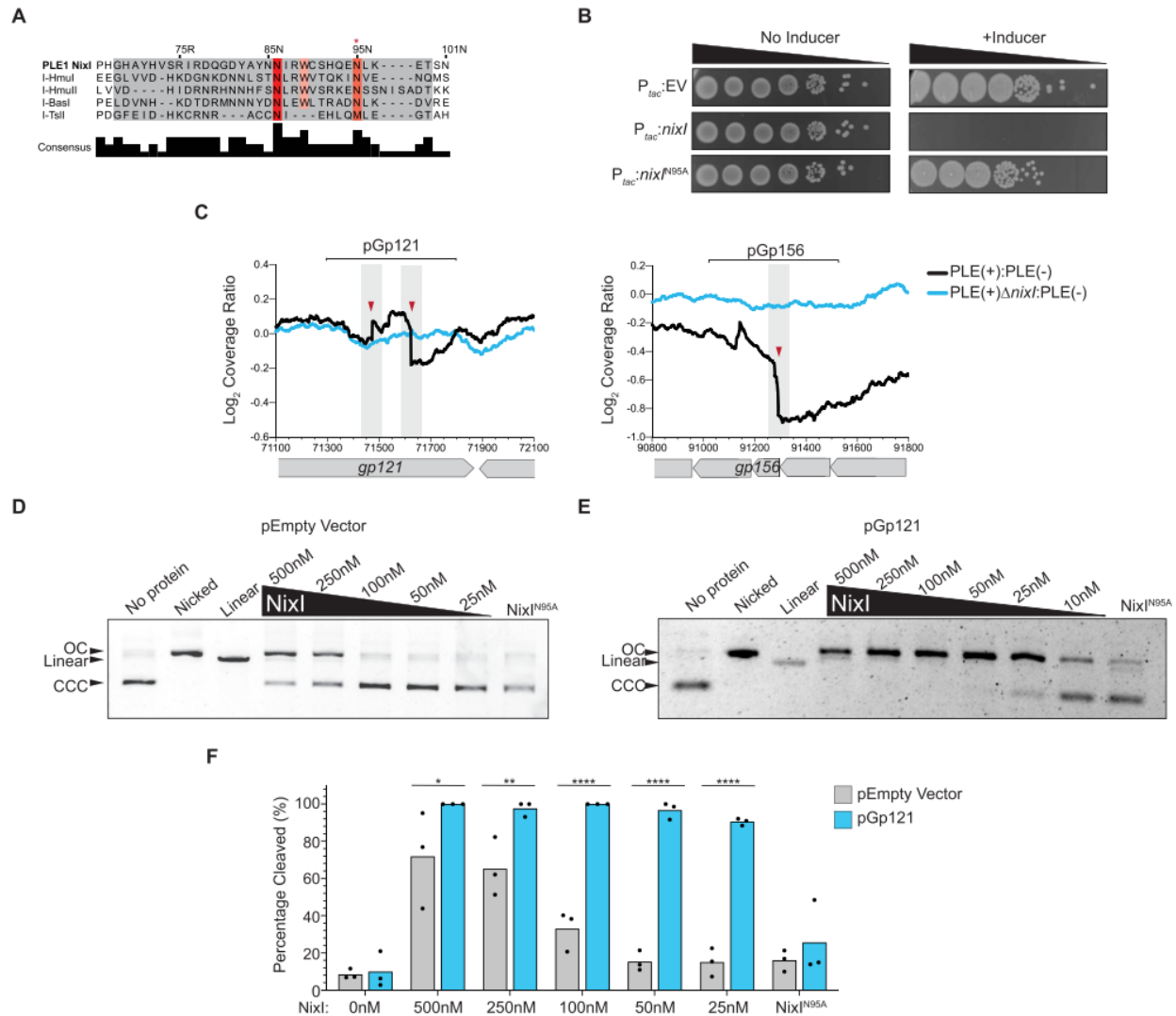


Figure 2.2. PLE encodes a nuclease, NixI, that cuts ICP1's genome in vivo and in vitro. (A) MUSCLE alignment of the catalytic core of phage-encoded intron homing endonucleases including I-Hmul, where I-Hmul's active site domain is highlighted in grey and residues that share >70% identity in orange/red. The red asterisk above 95N indicates the experimentally validated catalytic asparagine residue from I-Hmul (108). (B) 10-fold dilution series of mid-log grown *V. cholerae* containing plasmids for inducible expression of *nixI*, *nixI* with the putative catalytic asparagine mutated to an alanine (N95A), or an empty vector control, spotted on plates +/- inducer. (C) Coverage ratios PLE(+)/PLE(-) and PLE(+) $\Delta nixI$ /PLE(-) across ICP1's putative cleaved genes *gp121* and *gp156* as indicated in Figure 1. Brackets above indicate the regions cloned into plasmids and used as a DNA substrate for in vitro assays.

(D) *In vitro* cutting by purified NixI at the concentrations indicated above the gel or 500nM NixI^{N95A} of empty vector plasmid, showing linearized or nicked plasmid controls. (E) As in D, but showing cutting of pGp121. For D and E, CCC=covalently closed circular DNA, OC=open circular DNA. (F) Densitometry measurements calculating the percent of substrate cleaved at different concentrations of NixI (concentrations indicated on the x-axis) or 500nM NixI^{N95A} of the empty vector plasmid or pGp121. Bar height is the mean, dots are measurements from independent assays. Asterisks indicate statistically significant differences between cleavage of the empty vector plasmid and pGp121 (*P<0.0278, **P<0.0079, ****P<0.0001; where no asterisks are shown no difference was detected, two-way ANOVA, Sidak correction).

To directly evaluate NixI's nuclease activity, we purified NixI and NixI^{N95A} (Supplementary Figure 2.2) to conduct *in vitro* nuclease assays. Many nucleases require specific DNA topology such as negative supercoiling for activity, so we first assayed whether NixI had nucleolytic activity on plasmid substrates, allowing for visualization of both nicking, as evidenced by a loss of supercoiled plasmid, and double stranded breaks. We observed nuclease activity on an empty vector plasmid at high protein concentrations of NixI (≥ 250 nM) after a 15-minute incubation period (Figure 2.2D). We postulated that NixI exhibits a preference for cleavage of sites in ICP1's genome that we identified coverage drops in during infection of a PLE(+) host (Figure 2.2C). To test this, a 500 bp region flanking the putative cut sites in *gp121* was cloned onto the same plasmid and used as a substrate for NixI cleavage assays. NixI nucleolytic activity was dramatically enhanced for a plasmid template containing the region from *gp121*, allowing us to observe cleavage with only 25nM NixI (Figure 2.2E). Quantification of results from independent assays showed that cleavage of the plasmid harboring the region from *gp121* was significantly greater than cutting of an empty vector at all concentrations tested (Figure 2.2F). Importantly, nuclease activity was largely eliminated for NixI^{N95A} (Figures 2.2D-2.2F) even at a high protein concentration of 500nM, supporting our prediction from remote homology that NixI possesses an HNH/N catalytic motif (Figure 2.2A). To determine whether NixI is a nickase or double-stranded cutter, we compared cut products to plasmid controls that had been nicked or linearized using commercially available endonucleases with known sequence specificity. NixI generated one or more nicks in the DNA substrate but did not cause double stranded cuts (Figure 2.2D and 2.2E). We next assayed whether NixI cleaves PCR products. Even at high protein concentrations we observed little nucleolytic activity of PCR products, as visualized on native or denaturing agarose gels (Supplementary Figure 2.3). Together, these data show that PLE encodes a nicking endonuclease, NixI, that is necessary for cleavage of ICP1's genome *in vivo* and that shows *in vitro* specificity for sequences found in ICP1's genome. We propose that NixI is a PLE-encoded nuclease effector that targets ICP1, functioning as part of PLE's strategy to antagonize its viral host.

NixI's nuclease activity is dependent on the consensus motif GWAAWCTT

Having established that PLE-encoded NixI is a nickase with enhanced activity for a region from ICP1 that was cleaved *in vivo*, we sought to address if the consensus motif (Figure 2.1C) is necessary for NixI activity *in vitro*. We chose to focus on the region from *gp156* because it contains a single instance of the consensus motif while *gp121* contains two motifs. We made nucleotide substitutions to alter the motif from GTAATCTT to TGTATAGT, creating the plasmid pGp156*, which is identical to the plasmid harboring the region from *gp156* (pGp156) except for the changes to the 8 bp motif. We compared NixI's ability to cleave wildtype pGp156 and pGp156* in side-by-side nuclease assays and quantified the amount of substrate cleaved

(Figure 2.3A). NixI nuclease activity was significantly reduced for substrate with the mutated motif compared to the wildtype sequence at NixI concentrations of 25nM and 50nM (Figure 2.3A). We next asked whether this loss of nuclease activity could be explained by NixI's hampered ability to bind sequences lacking the consensus motif. We performed electrophoretic mobility shift assays (EMSAs) with probes internal to *gp156* that have the wildtype motif or the mutant derivative as in pGp156*. We found that DNA binding by NixI was dramatically reduced for substrate with the mutated motif (Figure 2.3B). Quantification of the amount of bound probe across replicate EMSAs showed a significant reduction in NixI binding to the probe with the mutated motif compared to the wildtype probe at all protein concentrations tested (Figure 2.3C). Together these data show that the motif GWAAWCTT is important for NixI DNA binding and cleavage.

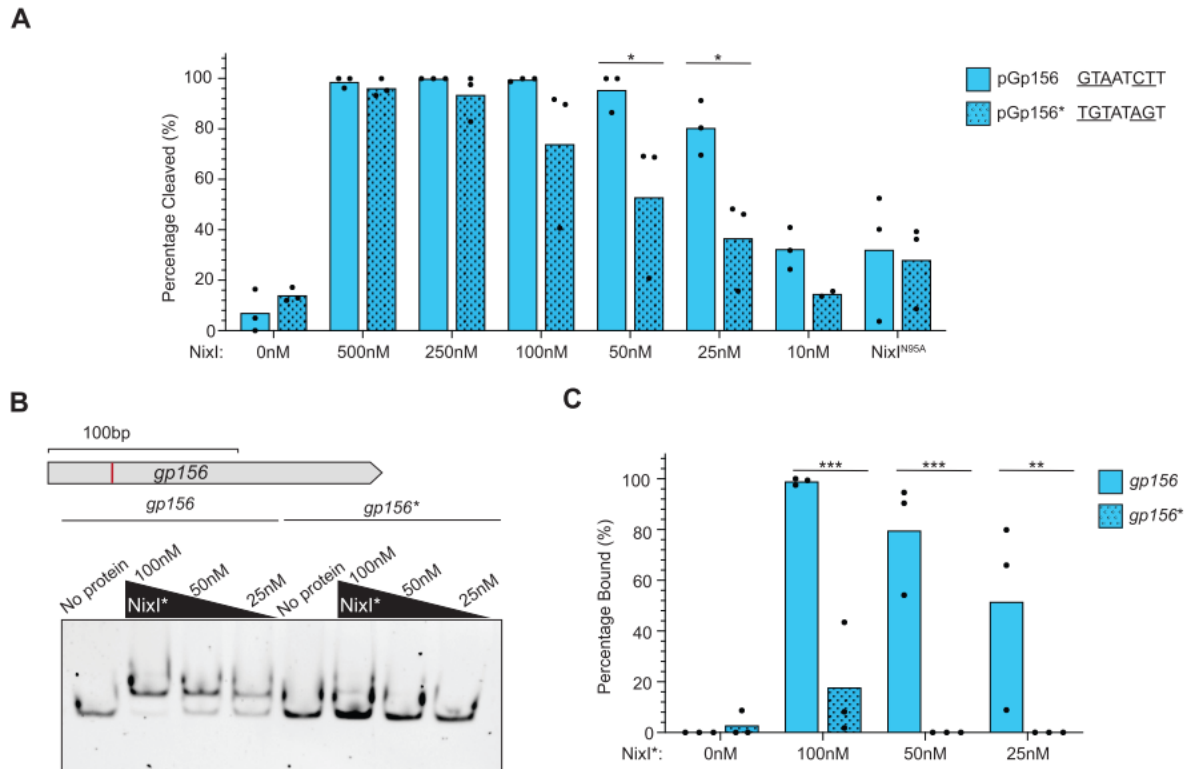


Figure 2.3. The consensus motif is required for NixI-mediated cutting and DNA binding (A) Densitometry measurements calculating the percent of substrate cleaved from *in vitro* nuclease assays at different concentrations of NixI (concentrations indicated on the x-axis) or 500nM NixI^{N95A} of pGp156 (as shown in Figure 2.2C) or pGp156*, which are identical vectors except for the nucleotide substitutions made in the consensus motif indicated to the right. Bar height is the mean, dots are measurements from independent assays. Asterisks indicate statistically significant differences between cleavage of pGp156 and pGp156* (*P<0.0332; where no asterisks are shown no difference was detected, two-way ANOVA, Sidak correction) (B) Electrophoretic mobility shift assays (EMSAs) using 100bp probes internal to *gp156* containing the consensus GWAAWCTT motif (top schematic, showing identified cut site in red) or the *gp156** mutant derivative incubated with NixI^{N95A} (NixI*) at the concentration indicated above the gel. (C) Densitometry measurements from three independent EMSAs showing NixI^{N95A} (NixI*) binding to the wildtype or mutated *gp156* probe. Bar height is the mean, dots are measurements from independent assays. Asterisks indicate statistically significant differences between binding of *gp156* and *gp156** (**P<0.0021, ***P<0.0002; where no asterisks are shown no difference was detected, two-tailed t-test, Sidak correction).

NixI inhibits ICP1 progeny production and replication

Having shown that NixI is necessary and sufficient to generate the cleavage patterns observed in ICP1's genome during infection of a PLE(+) host, we next asked whether NixI activity inhibits ICP1's lifecycle. Previous work has identified PLE-encoded mechanisms that effect ICP1 gene expression (35) and that collapse ICP1-induced lysis inhibition (100), however, none of these mechanisms alone are sufficient to prevent ICP1 progeny production, showing that there are additional mechanisms of PLE-mediated ICP1 interference. We anticipated that NixI cleavage of ICP1's genome would have negative consequences for ICP1's capacity to produce phage progeny. Because of NixI's toxicity to *V. cholerae* (Figure 2.2B) we were unable to assess plaque formation in a strain expressing NixI. Therefore, we assayed *nixI*-mediated inhibition with an efficiency of center of infection assay, which measures ICP1 virions produced from a single round of infection of *V. cholerae* by collecting virions produced from an infected host of interest and quantifying the progeny phage produced on a permissive host using a plaque assay (Figure 2.4A). *V. cholerae* expressing *nixI*, but not *nixI*^{N95A}, reduced ICP1 progeny production 100-fold (Figure 2.4B). However, *nixI* was not required for PLE-mediated inhibition of ICP1, as PLE(+)Δ*nixI* was still inhibitory to ICP1 (Figure 2.4C). This corroborates previous work that showed no single PLE-encoded ORF, including *nixI*, is necessary for inhibition of ICP1 by plaque assay (100). NixI, however, is sufficient to limit ICP1 progeny production in a single round of infection, the first PLE-encoded product shown to recapitulate the block in progeny production from wildtype PLE (Figure 2.4B and 2.4C).

We anticipated that *nixI* inhibits ICP1 progeny production by impairing phage genome replication, which is a hallmark of PLE activity (101). We therefore measured ICP1 genome replication by qPCR in a wildtype PLE(+) host compared to a Δ*nixI* mutant derivative. As was previously observed (101), PLE reduced ICP1 genome copy to ~20% of the levels observed in a permissive infection. Deletion of *nixI* restored ICP1 replication to ~65% of the levels of replication observed in a permissive host (Figure 2.4D). Complementation of *nixI* *in trans* reduced ICP1 replication to a greater extent than wildtype PLE (Figure 2.4D). This is likely because endogenous *nixI* is not expressed until mid-way through infection, between 8 and 12 minutes post-infection (59). At this point in infection, ICP1's has undergone several rounds of theta replication and is transitioning to rolling circle replication, while PLE replication begins to accelerate (101). Pre-induction of *nixI* as was done here (Figure 2.4D) would likely interfere with all stages of ICP1 replication and could therefore exceed PLE's impact on ICP1 DNA replication. To assess if NixI directly impacts ICP1's genome replication in the absence of other PLE-encoded products, we also evaluated ICP1 genome replication in a PLE(-) background. As seen with *nixI* complementation *in trans*, ectopic expression of *nixI* was sufficient to inhibit ICP1 replication to 10% that of a permissive infection, again greater inhibition than is observed in the context of wildtype PLE (Figure 2.4E). Induction of the catalytically dead *nixI*^{N95A} did not inhibit ICP1 replication (Figure 2.4E), demonstrating that catalytic activity of NixI is necessary to inhibit ICP1 replication. Together these data demonstrate that *nixI* is a potent inhibitor of ICP1 replication, resulting in decreased virion production.

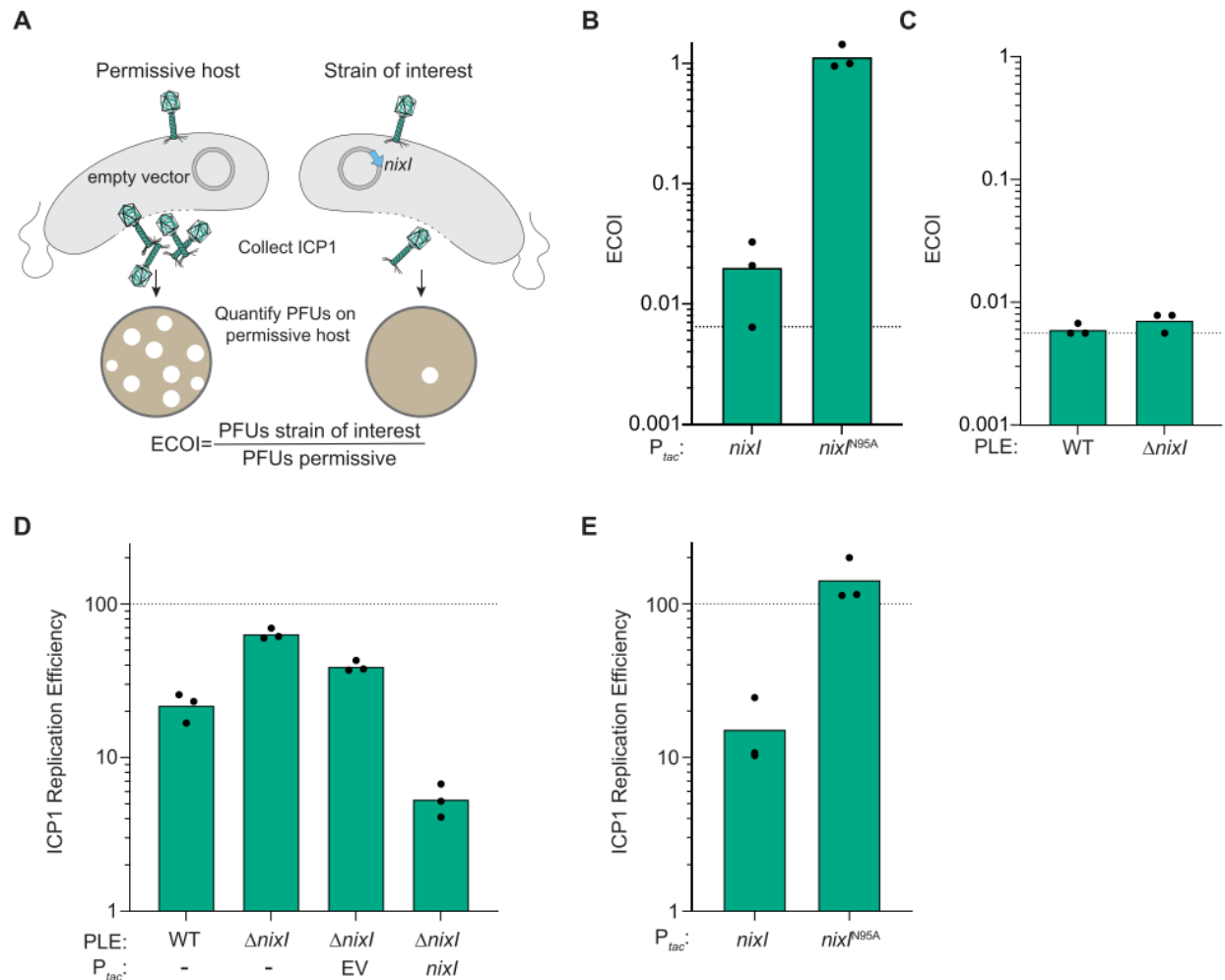


Figure 2.4. NixI is necessary and sufficient to inhibit ICP1 genome replication and sufficient to inhibit production of ICP1 progeny (A) Schematic of efficiency of center of infection (ECOI) assay, where ICP1 produced by a single round of infection in a strain of interest (here, expressing *nixI*, on the right) is quantified as a ratio of the number of phage produced by a permissive host (left). (B) ECOI of ICP1 on *PLE(-)* *V. cholerae* expressing *nixI* or *nixI^{N95A}* from a plasmid, relative to phage progeny produced from an induced empty vector control. The dashed line indicates where the progeny phage output is equal to the input phage. (C) ECOI of ICP1 in *PLE(+)* *V. cholerae* or *PLE(+)* $\Delta nixI$ relative to phage progeny produced from a permissive control. The dashed line indicates where the progeny phage output is equal to the input phage. (D) ICP1 replication efficiency in the *PLE(+)* strains indicated relative to infection of permissive *PLE(-)* *V. cholerae* harboring an empty vector control (dashed line) 20 minutes post-infection as assessed by qPCR. (E) ICP1 replication efficiency in *PLE(-)* *V. cholerae* pre-induced to express the gene indicated relative to infection of permissive *PLE(-)* *V. cholerae* harboring an empty vector control (dashed line) 20 minutes post-infection as assessed by qPCR. For B-E bar height displays mean and dots indicate independent biological replicates.

PLE is protected from cleavage through a sequence-independent mechanism

Organisms that use nuclease effectors as weapons for inter-genome conflict must ensure that the nuclease can discriminate between self and non-self DNA. We have shown that PLE encodes a nickase that cleaves ICP1's genome leading to a decrease in replication and phage progeny production. Despite NixI's ability to restrict phage genome replication, PLE replicates

to high copy number and transduces to new hosts, suggesting PLE is resistant to NixI activity. Indeed genome integrity is important for PLE's lifecycle since ICP1 can antagonize PLE by employing its own nuclease effectors, namely Cas2-3 and Odn, which limit PLE replication and transduction (34, 102). We therefore anticipated that PLE would be protected from NixI-mediated cleavage rather than sacrifice its own genome to thwart ICP1.

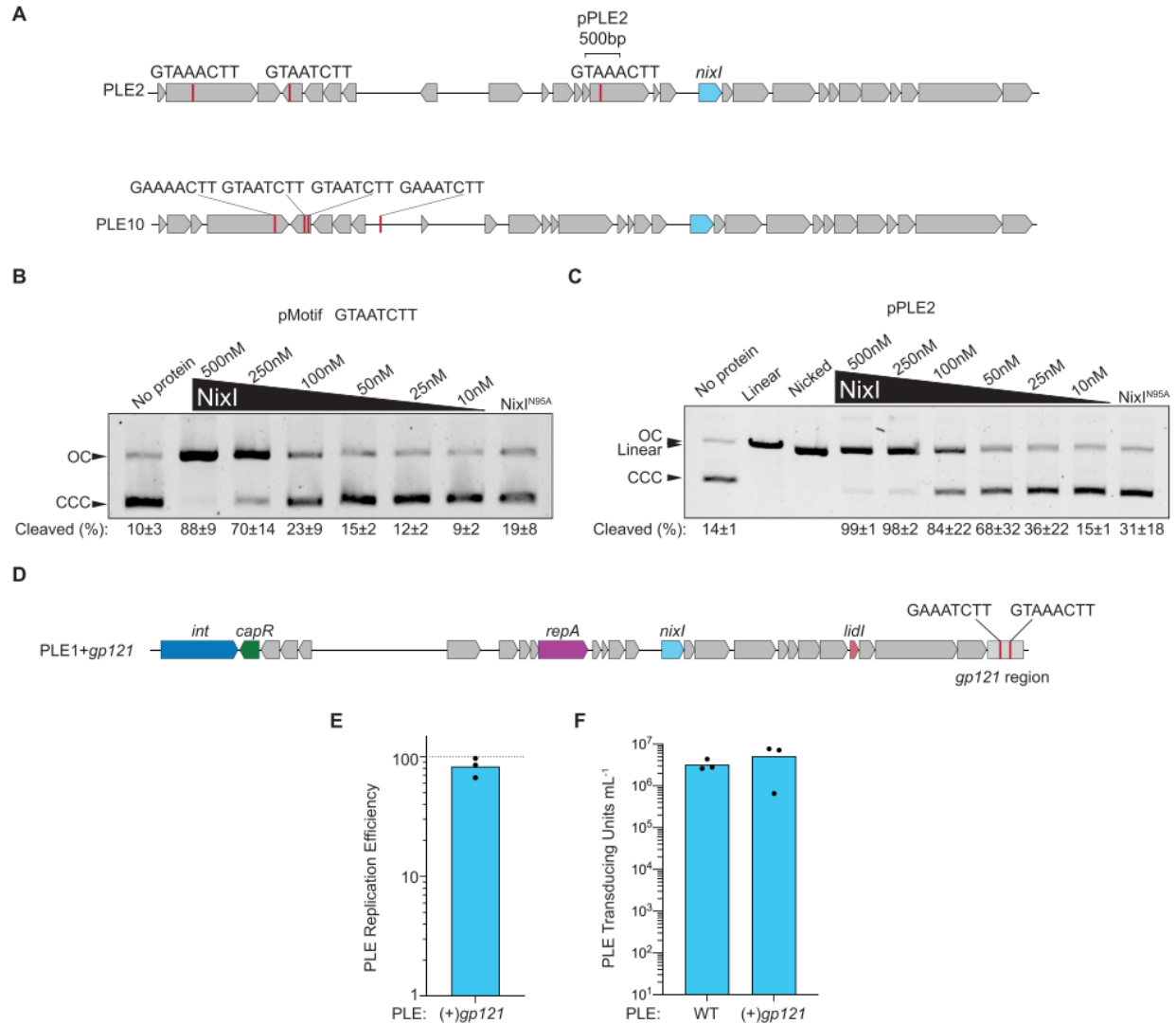


Figure 2.5. PLE is protected from NixI-mediated cleavage independent of sequence identity. (A) Gene maps of PLE2 and PLE10 showing instances of the consensus motif as red lines and the region cloned to create pPLE2, and the gene that encodes for identical amino acid identical NixI in blue. (B) *In vitro* cutting by NixI (concentrations indicated above the gel) or 500nM NixI^{N95A} of pMotif which contains only the 8bp motif from *gp156* (indicated above the gel) displaying the percent of cleaved substrate as the average of three independent assays and the standard deviation below the gel. CCC=covalently closed circular DNA, OC=open circular DNA. (C) As in B, but showing cutting of pPLE2 and displaying linearized and nicked plasmid controls. (D) Gene map of engineered PLE1 where the region from ICP1's *gp121* shown to be cut *in vivo* and *in vitro* was cloned at the 3' end of the integrated PLE. (E) Replication efficiency of PLE1 possessing ICP1's *gp121* region relative to the wildtype PLE (dashed line) 20 minutes post-infection as assessed by qPCR. (F) Transduction of wildtype PLE1 and PLE1 with the *gp121* region. For E and F bar height is the mean, dots are measurements from independent assays.

To address how PLE is protected from NixI activity, we first asked whether PLE contains the consensus motif, which we demonstrated is an important determinant of NixI activity (Figure

2.3). We found that PLE1, which has been the focus of our study, does not contain the consensus motif, suggesting that PLE may simply avoid NixI-mediated cleavage based on sequence alone. To corroborate this finding, we next looked at whether PLEs encoding identical NixI proteins also lack the consensus motif. Surprisingly, both PLE2 and PLE10, which have amino acid identical copies of PLE1's NixI, contain three and four instances of the motif, respectively (Figure 2.5A). This prompted us to test whether the motif alone is sufficient to direct NixI-mediated cutting. We cloned the 8bp motif GTAATCTT from *gp156* into the plasmid (generating pMotif) and assessed cleavage of this substrate *in vitro*. Cutting of pMotif was greatly reduced compared to pGp156 which has additional sequence flanking the motif. For example, only 15% of pMotif was nicked compared to nearly 95% of pGp156 at 50nM of NixI, (Figures 2.3A and 2.5B). Overall, cleavage of pMotif was comparable to cutting of the empty vector at all concentrations tested (Figures 2F and 5B). This demonstrates that although the GWAAWCTT motif is necessary for NixI cleavage and binding (Figure 2.3), it is not sufficient to direct NixI cutting, suggesting that additional unidentified sequence features flanking the motif may be required.

We next asked whether regions from PLEs that contain the motif avoid NixI activity because they lack additional sequence features required for cutting. Towards this, we cloned a 500bp region flanking one of the identified motifs from PLE2 into the plasmid (referred to as pPLE2, Figure 2.5A) and assayed NixI cutting *in vitro* (Figure 2.5C). Cleavage of pPLE2 approximated cleavage of pGp156 at protein concentrations $\geq 50\text{nM}$, suggesting that this region from PLE2 has the sequence features that are sufficient for NixI cutting *in vitro*. Having determined that regions from PLE2 would not be protected from NixI activity due to sequence identity alone, we sought to reconcile this finding with previous studies showing that PLE2 achieves robust replication and transduction (32) and expresses NixI during ICP1 infection (59). We therefore asked whether we could make PLE1 sensitive to NixI activity by adding to it the region from *gp121* that contains two motifs and exhibited robust cleavage *in vivo* and *in vitro* (Figure 2.2), predicting PLE1 with the *gp121* region ('PLE+*gp121*', Figure 2.5D) would suffer a replication defect as we observed for ICP1. Strikingly, PLE harboring the cut region from *gp121* was able to replicate as well as wildtype PLE, as there was no significant difference between PLE replication with and without the cut sites ($P=0.1767$, paired t test) (Figure 2.5E). Given that qPCR to assess PLE replication amplifies only a small region of the genome which is 4kb away from the introduced *gp121* sites, we also evaluated if introduction of the *gp121* region manifested as a transduction defect. Consistent with our finding of no evidence of a replication defect, PLE+*gp121* transduced as well as wildtype PLE (Figure 2.5F). These data show that PLE is protected from NixI cutting even when it harbors sequence that is robustly cut *in vitro* and cut in the context of ICP1's genome *in vivo*. Collectively these data show that PLEs have not evolved to avoid NixI activity on the basis of sequence alone, but suggest that PLEs are protected from NixI-mediated cleavage through a yet unidentified mechanism.

NixI promotes PLE replication

PLE replication relies on hijacking replication resources from ICP1, leading us to hypothesize that PLE may limit ICP1 replication to successfully co-opt replication machinery for its own replication. This suggested that NixI-mediated cleavage of ICP1's genome and its subsequent limitation of ICP1 replication may be necessary for PLE to replicate efficiently. We therefore assayed PLE replication efficiency of PLE $\Delta nixI$ relative to wildtype PLE by qPCR

(Figure 2.6A). During ICP1 infection we observed that PLE replication is reduced to 25% of wildtype in absence of *nixI* (Figure 2.6A). Complementation of *nixI* *in trans* restored PLE replication to wildtype levels. Importantly, these data help us to contextualize the effect of *nixI* expression on ICP1 replication. Because *nixI* induction is toxic to *V. cholerae* (Figure 2.2B), we initially had concerns that the observed reduction in ICP1 replication in the presence of *nixI* (Figure 2.4E) was the result of indirect effects on *V. cholerae*, rendering the host unable to support phage infection and DNA replication in general. However, PLE replication is increased by *nixI* expression *in trans* (Figure 2.6A), demonstrating that *nixI*-mediated inhibition of ICP1 replication is specific and not a pleiotropic effect of cellular toxicity.

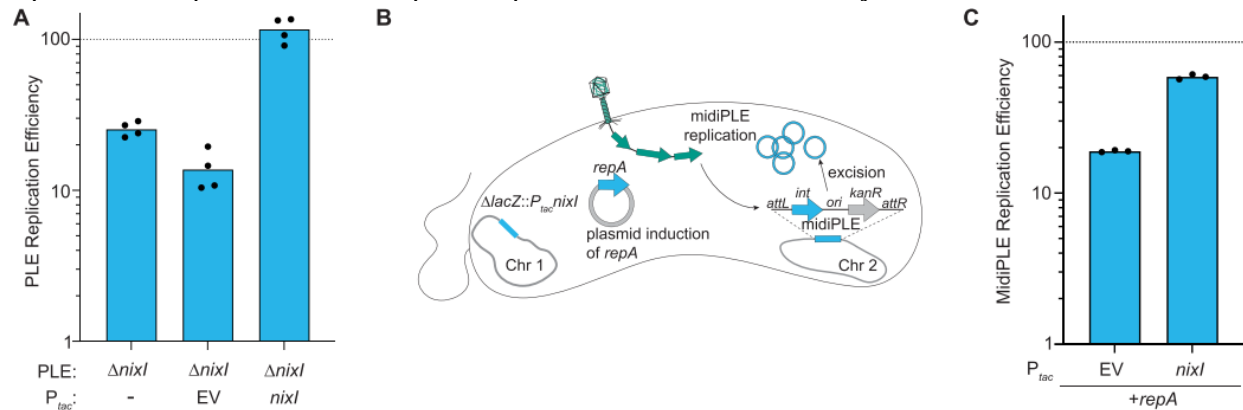


Figure 2.6. Expression of *nixI* is required for PLE replication and can augment replication of a minimal replicating PLE (midiPLE) (A) Replication efficiency of *V. cholerae* PLE mutant and complemented derivatives 20 minutes post-infection by ICP1 relative to replication of wildtype PLE (dashed line). (B) Model of midiPLE induction and replication during ICP1 infection. Plasmid-encoded *repA*, PLE's essential replication protein (101) and chromosomally encoded *nixI* or an empty vector were pre-induced for 15 minutes prior to infection by ICP1. ICP1 infection triggers midiPLE excision from the *V. cholerae* chromosome and replication of midiPLE is detected by qPCR. (C) Replication efficiency of midiPLE in the presence and absence of *nixI* relative to wildtype PLE (dashed line) 20 minutes post-infection. For A and C, bar height indicates mean, and dots are independent biological replicates.

Previous work characterizing PLE replication identified PLE's minimal unit of replication, (referred to as midiPLE) comprised of PLE's attachment sites, integrase, and origin of replication integrated into PLE's native locus in *V. cholerae* (Figure 2.6B). When ICP1 infects, midiPLE excises from the chromosome, and in the presence of PLE's essential replication initiation factor, *repA*, midiPLE replicates to high levels; however, midiPLE replication does not reach the same robust levels of replication as is observed with the full length wildtype PLE (101). As demonstrated above, NixI is required for wildtype levels of PLE replication (Figure 2.6A), therefore we asked whether NixI could boost midiPLE replication to levels more comparable to wildtype PLE. Induction of *nixI* increased midiPLE replication to ~60% of wildtype PLE replication, greater than *repA* driven midiPLE replication alone (Figure 2.6C). The impact of NixI on midiPLE replication could be due to inhibition of ICP1, or NixI may have a direct role in PLE replication, as nicking enzymes are required factors for initiating RCR of plasmids and viral genomes (103). Previous work characterizing PLE's replication profiles over the course of infection suggested that PLE also undergoes a transition to RCR (101), which we anticipate would be necessary for PLE to linearize its genome for packaging. We postulated that NixI may act as the nicking nuclease for initiating PLE's RCR, which is often the same region where packaging is initiated from (the *pac* site). To test whether NixI can cleave PLE's RCR ori, we identified a candidate region in PLE's ORF-less space to clone into our

plasmid as a substrate for nuclease assays. We reasoned that PLE's RCR ori would be highly conserved among all known PLEs, and cloned a 500bp region previously identified as highly conserved (102) (Supplementary Figure 2.4A). We found that even at high protein concentrations NixI exhibited minimal cleavage of this substrate suggesting NixI does not directly cut PLE's RCR ori (Supplementary Figure 2.4B). Additionally, if NixI directly cuts midPLE's RCR origin, pre-induction of *nixI* would be detrimental to midPLE replication, as it would nick the origin potentially before the expression of the remaining proteins required for RCR (such as single-stranded DNA-binding protein) and midPLE would likely fail to replicate. Because NixI increases midPLE replication (Figure 2.6C), even when ectopically expressed prior to infection, we anticipate NixI is not PLE's RCR initiator nickase, but rather cleaves ICP1's genome to liberate replication resources for midPLE. The targeted nicking of ICP1 could liberate replication resources for PLE and hinder ICP1's ability to package full-length genomes.

Homologs of NixI are a conserved feature of an expanded family of phage satellites in *Vibrios*

To date, PLEs have only been found in toxigenic *V. cholerae* and have been shown to be dependent on ICP1. Accordingly, all PLE(+) *V. cholerae* that have been isolated possess the lipopolysaccharide O1 antigen that serves as the receptor for ICP1 (4). Bioinformatic searches for PLEs in sequenced genomes have searched for nucleotide similarity, as PLEs 1-10 all share homology at the nucleotide level. In an effort to predict NixI's domain organization and function, we used HHblits (110) to look for homologs of NixI from PLE1 and found proteins that share ~25% amino acid identity in non-toxigenic *V. cholerae* (e-value=0.0000072). Notably, NixI homologs outside of those found in PLEs in toxigenic *V. cholerae* do not share detectable homology at the nucleotide level. Further investigation of the genomic context of these NixI homologs identified an expanded family of putative phage parasites similar to PLEs, as we elaborate on below. To determine whether the satellites from non-toxigenic *V. cholerae* were indicative of an even larger family of parasites, PSI-BLAST searches were performed with the NixI homologs from putative satellites in non-toxigenic *V. cholerae* and excluding *V. cholerae* as a subject. This search returned several homologs in other *Vibrios* such as *V. parahaemolyticus* and *V. casei*. To expand this family of potential satellites, reciprocal searches were performed with the tyrosine recombinase/integrase from *V. parahaemolyticus* strain s042 and additional elements were gathered. In all cases, the NixI homologs were found on putative mobile genetic elements which possess key features of PLEs. First, they all encode a 5' integrase/recombinase; either a large serine recombinase (which share ~50% amino acid identity with PLE-encoded large serine recombinases from toxigenic *V. cholerae*) or a tyrosine recombinase (Figure 2.7A). All putative parasites are ~17kB in length and, like PLEs from toxigenic *V. cholerae*, encode for almost exclusively hypothetical proteins. However, PFAM domain searches indicated that the satellites have ORFs with domains in previously characterized PLE-encoded ORFs (Figure 2.7A). Specifically, we searched for proteins with homology or shared domains with PLE1's CapR (T5ORF172 domain) (35), RepA (RepA_N domain) (101) and LidI (transmembrane domain) (100). All putative satellites encode a predicted replication initiation protein at the same location as PLEs' RepA, but interestingly, we saw two different replication proteins: either a RepA_N domain or a DNA Pol I domain. The putative parasites have a small transmembrane domain protein similar to PLE's LidI (100), suggesting that disruption of lysis inhibition may be a conserved function of phage parasites found in other *Vibrio* species. None of the parasites

possessed a protein with a T5ORF172-domain protein. In addition, all of the putative parasites are syntenic with PLEs, possessing a large ORF-less space where PLEs' origin of replication is located (101), as well as a smaller ORF-less space upstream of *nixI*, where PLEs encode a predicted small RNA (59). Although this analysis is not exhaustive, it shows for the first time that putative parasites similar to PLEs are found outside of toxigenic *V. cholerae*.

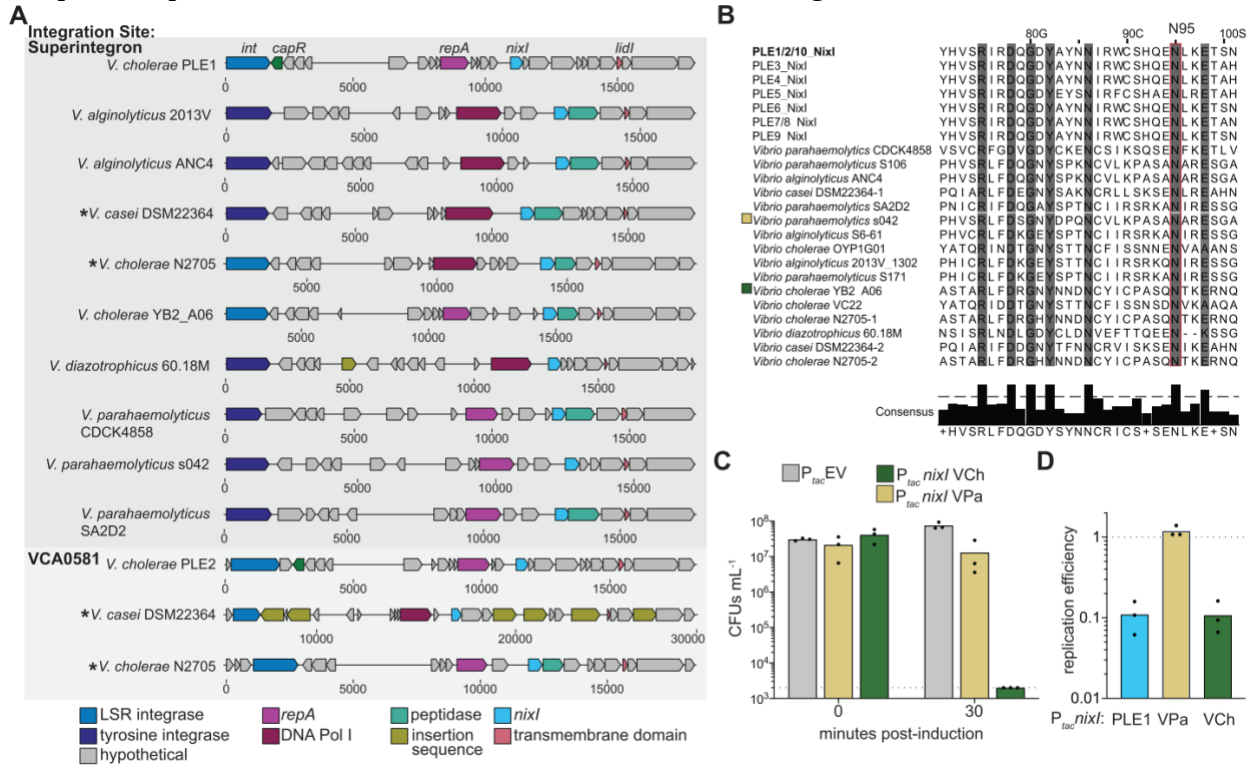


Figure 2.7. NixI homologs characterize an expanded family of putative phage parasites in *Vibrio* species (A) Gene maps (scale below map displays relative nucleotide position) of putative phage parasites from *Vibrios*, where homologs of functionally characterized genes from PLE1 (labeled top) are colored according to shared predicted function identified by PFAM domains or HHPred. Top: phage parasites integrated into the repeat regions of the superintegron, bottom: phage satellites integrated into *vca0581*, the integration site shared by PLEs 2, 7, 8 and 9. Asterisks are used to note where two similar phage satellites occur in different locations in the same genome. LSR = large serine recombinase. (B) MUSCLE alignment of NixI homologs, showing the experimentally determined catalytic asparagine from PLE1's NixI (identical to NixI from PLE's 2 and 10) boxed in red (N95), and residues sharing >70% consensus shaded in grey. When more than one satellite was found in the same strain, they are differentiated by "-1" or "-2" after the strain name. (C) Cell viability as measured by CFUs ml⁻¹ of *V. cholerae* expressing *nixI* homologs from *V. cholerae* YB2_A06 (VCh) and *V. parahaemolyticus* s042 (VPa), plated prior to induction and 30 minutes post-induction. The dashed line indicates the limit of detection. (D) Same colors as in C, showing ICP1 replication efficiency in strains pre-induced to express *nixI* homologs relative to *V. cholerae* harboring an empty vector control (dashed line) 20 minutes post-infection as assessed by qPCR. For C and D, bar height displays mean and dots indicate independent biological replicates.

We next looked at the sequence flanking the putative satellites to determine if they integrate into conserved attachment sites in their bacterial hosts. In the laboratory, PLEs 1, 3, 4, 5, 6 and 10 from toxigenic *V. cholerae* integrate into *V. cholerae* repeats (VCRs) that occur >100 times in the superintegron, each serving as a potential attachment site (32, 33). Although long-read sequencing is required to resolve the repetitive nature of the superintegron to definitively state where these putative satellites integrate, the majority are integrated into the superintegron repeats even though some possess a tyrosine recombinase unrelated to the large serine

recombinases found in PLEs from toxigenic *V. cholerae*. PLEs 2, 7, 8 and 9 integrate into the M48 metallopeptidase gene (*vca0581*) outside of the superintegron (32, 33). Interestingly, two of the identified *Vibrio* genomes, *V. cholerae* N2705 and *V. casei* DSM22364, have multiple putative satellites: one integrated into a superintegron repeat and a second in the M48 metallopeptidase gene. Multiple PLEs in a single *V. cholerae* genome has not previously been described. The presence of multiple satellites similar to PLEs found in the same genome suggests the modular nature of PLEs that has been observed (32, 33) is driven by recombination between satellites in the same host. The second putative satellite in *V. casei* (in the M48 metallopeptidase gene) is littered with insertion sequences, suggesting that there may be no fitness benefit to having multiple phage satellites and that they can be lost through degeneration (Figure 2.7A). NixI homologs from PLEs in toxigenic *V. cholerae* are all highly similar, sharing >65% amino acid identity. In contrast, NixI homologs from the putative satellites outside of toxigenic *V. cholerae* are highly divergent, sharing only 20-25% amino acid identity (Supplementary Table 2.3). A MUSCLE alignment of all unique NixI homologs from *Vibrio* species indicates that the HNH/N catalytic asparagine identified in this study is conserved (Figure 2.7B). A neighbor-joining tree based off this alignment shows NixI encoded by PLEs from toxigenic *V. cholerae* are closely related, while the NixI homologs from non-toxigenic *V. cholerae* cluster together (Supplemental Supplementary Figure 5). NixI from other *Vibrios* did not group by species, suggesting the potential for genetic exchange between putative satellites found in the genomes of *V. parahaemolyticus*, *V. alginolyticus* and *V. casei* (Supplementary Figure 2.5). To determine if NixI is functionally conserved among the newly identified putative phage satellites, we expressed *nixI* from *V. cholerae* YB2_A06 (an environmental non-toxigenic isolate) and from *V. parahaemolyticus* s042 in our lab strain of toxigenic *V. cholerae*, and the impact of NixI induction on cell viability was assayed. Induction of the homolog from *V. cholerae* YB2_A06 completely inhibited colony formation of our lab strain of *V. cholerae*, while the homolog from *V. parahaemolyticus* was only mildly toxic (Figure 2.7C). To determine if these *nixI* homologs are inhibitory to ICP1, we quantified ICP1 replication in *V. cholerae* pre-induced to express either homolog, or *nixI* from PLE1, and compared replication efficiency to that of a permissive host with the empty vector. Interestingly, the homolog from *V. cholerae* YB2_A06 inhibited ICP1 replication, yet the homolog from *V. parahaemolyticus* did not (Figure 2.7D). *V. cholerae* YB2_A06 is a non-O1 isolate that would not be susceptible to ICP1 infection. Induction of PLE's genetic program requires infection by ICP1 (59), so specificity of NixI homologs for their cognate phages could be driven by expression only in response to infection by their viral host, rather than recognition of a motif unique to their host phage's genome. Together, we show that PLEs are not restricted to toxigenic *V. cholerae*, but functionally analogous elements also reside in non-toxigenic *V. cholerae* as well as other *Vibrios*. Although members of this expanded family of phage satellites remain to be experimentally validated, these data suggest that PLE-encoded strategies that restrict ICP1, such as *nixI*-mediated cleavage, may be conserved among divergent satellites and represent a mechanism to parasitize a broader class of inducing phage.

Discussion

Phage parasites can be considered host defense systems, as the parasite inhibits production of its viral host, benefiting the bacterial population (32, 91). However, phage parasites are fundamentally selfish elements, promoting both their *vertical* transmission by protecting the bacterial host population from viral takeover, as well as their own *horizontal* transmission

through parasitism of their viral host's structural components (35, 94, 111). Phage parasites employ finely tuned mechanisms to prevent packaging of the viral host's genome, but allow for progression of the viral host's gene expression program for production of structural proteins which are hijacked by the phage parasite. For this reason, it is perhaps unsurprising that all phage parasites studied outside of *V. cholerae* reduce but do not eliminate production of their viral host (94, 96, 112, 113). *V. cholerae*'s PLEs are unique in their capacity to completely restrict their viral host ICP1 (32), yet, the PLE-encoded mechanisms responsible for ICP1 restriction have remained mysterious. In this work, we show that NixI from PLE1 functions as a nuclease effector, adopting a strategy that is widely used by bacterial defense systems to restrict phage production (7, 8, 10). We propose a model where NixI functions as a nicking endonuclease, introducing site-specific nicks into ICP1's genome. Anti-phage defense systems that encode nicking endonucleases have recently been characterized, suggesting nicking is an effective strategy to destroy invading viral DNA (114, 115). Although we did not detect any exonuclease activity from NixI *in vitro*, the coverage profiles across ICP1 suggest that additional exonuclease(s) may act 3' of the cleaved regions, perhaps as an effort by ICP1 to repair DNA damage. While nicking may directly cause replisome disassembly (116), attempts to repair nicks may also recruit DNA repair and recombination factors to these sites, stealing them away from collapsed replication forks leading to further defects in replication (117). We show here that NixI is the most inhibitory PLE-encoded product characterized thus far, recapitulating PLE's inhibition of ICP1 progeny production and reducing ICP1 replication by an order of magnitude. Targeting viral host replication with a nuclease effector is a strategy of viral host interference that, to our knowledge, has not been previously reported for phage parasites. Though as we discuss below, PLE-mediated restriction of ICP1 replication through NixI does not completely prevent ICP1 replication, as PLE relies on the maintenance of ICP1's genome for expression of replication (118) and structural (35) proteins that PLE parasitizes. NixI is not necessary for PLE-mediated inhibition of ICP1, consistent with there being multiple PLE-encoded mechanisms to restrict ICP1 (100). While the remaining mechanisms of ICP1 restriction remain to be determined, they may target nodes in ICP1's lifecycle to divert structural components towards PLE packaging, a strategy used by all known phage parasites studied to date (98, 119, 120).

Several lines of evidence implicate NixI as the PLE-encoded factor that inhibits ICP1's transition to RCR. First, bioinformatic analyses identified NixI as the only putative nuclease encoded by PLE. Additionally, transcription of the *nixI* operon peaks between 12 and 16 minutes post-infection, at a time when ICP1 is transitioning to RCR (59, 101). NixI is necessary for the appearance of cut sites in ICP1's genome, which are evident at 16 minutes post-infection of PLE(+) *V. cholerae*. Inhibiting ICP1 replication may liberate ICP1- and/or *V. cholerae*-encoded replication resources for PLE. The components of ICP1's and PLE's replisomes remain to be experimentally validated, though ICP1 is predicted to have a T7-like replisome comprised of DNA Pol-I, a primase-helicase, an RNaseH homolog and a yet unidentified single-stranded DNA-binding protein (17, 121). PLE lacks homologs of these key replisome components, suggesting that it hijacks ICP1's replisome, which is supported by experimental evidence showing that PLE replication depends on an ICP1-encoded helicase (118). PLE-mediated inhibition of ICP1's RCR would prevent ICP1 from linearizing its genome, which is necessary for efficient packaging. In this way, inhibition of RCR may be functionally analogous to a strategy used by phage parasites in *Staphylococcus aureus* (known as SaPIs) to prevent packaging of their viral host's genome. SaPIs inhibit their viral hosts' small terminase (TerS) subunit and encode their own replacement (122). TerS is the component of the packaging

machinery that specifically recognizes the viral DNA to initiate translocation of the viral genome into virions (106). By blocking TerS and encoding their own, SaPIs ensure that their own genome is preferentially packaged. Adopting a similar strategy, divergent satellites found in *E. coli* encode a protein that redirects the viral host's TerS to specifically package the parasites' genome (123). Our analyses suggest that ICP1 packaging is still inhibited by PLE Δ nixI, as cleavage of ICP1's *pac* site is still blocked (Supplemental Supplementary Figure 1B). This suggests that PLE encodes additional inhibitory mechanism(s) that may block ICP1's terminase complex, and may act synergistically with NixI to ensure ICP1's genome is not packaged.

An essential component of nuclease effectors is the ability to specifically target foreign DNA. We therefore anticipated that PLE would have mechanisms to protect itself from self-targeting by NixI. Based on the work presented here, we hypothesize that PLE is protected through other mechanisms that are more nuanced than distinguishing self from non-self based on sequence alone. Recent work has identified an anti-CRISPR protein that counters CRISPR-Cas immunity by nicking supercoiled DNA, which is the preferred substrate for the SpyCas9 nuclease effector, allowing for phage escape from CRISPR targeting (124). Here we have shown that negative supercoiling, which is the state of plasmid substrates used for *in vitro* cutting assays, is robustly cleaved by NixI while we saw no evidence that linear PCR products are nicked, suggesting NixI activity depends on DNA topology. It is therefore possible that PLE possesses alternative DNA topology to avoid NixI cleavage, perhaps by alleviating its own supercoiling, while NixI may preferentially cleave supercoiled regions of ICP1's genome. We cannot rule out the possibility that PLE might not be protected from NixI cleavage, but rather is able to overcome NixI targeting through robust genome replication. PLE replicates to ~1000 genome copies per cell, exceeding the copy number increase of ICP1. PLE possesses multiple mechanisms to hinder ICP1's genome replication, as ICP1 replication is still reduced in a PLE(+) Δ nixI background. This defect in replication may render ICP1 more susceptible to NixI activity, while robust PLE replication may outpace the rare nicks introduced by NixI on a self-targeting PLE. We further observe many instances of NixI-dependent cleavage drops in ICP1's genome *in vivo* during infection of a PLE(+) host, suggesting that cutting at many regions is required for NixI to hamper ICP1 replication. However, if PLE is not protected from NixI activity, PLE Δ nixI should show elevated replication. Rather, we observe a replication defect in the absence of *nixI*. However, it is possible that there may be counter-acting forces at play: NixI cleaves ICP1's genome limiting competition with its viral host and liberating replication resources for PLE, yet NixI can nick PLE hampering its replication. These additive effects may result in the intermediate PLE replication defect seen with PLE Δ nixI. Determining how PLE is protected from self-targeting may reveal new ways of self-non-self discrimination by nuclease effectors.

Established and novel phage defense systems can be predicted *in silico* through the presence of conserved domains or by their genomic proximity to known defense systems (44, 45). Our work complements recent studies showing that phage parasites are likely ubiquitous in bacterial genomes (98, 99) and highly similar parasites can be found in divergent taxa, such as the parasite P4 that was previously thought to be restricted to *Escherichia coli* but was recently found across several genera of Enterobacterales (125). Phage parasites lack hallmark genes allowing for easy identification but can be identified as mobile genetic elements carrying an integrase/recombinase. Additional gene cargo carried by phage parasites can influence the fitness of their bacterial hosts, beyond their established role in restricting viral host propagation, as phage parasites have recently been shown to encode defense systems that function more broadly

against infecting phages (126). Demonstrating bona fide phage parasitism is challenging given that a compatible tripartite system must be identified. While the cellular host is known from the integrated parasite genome, the cognate viral host may not be as obvious, especially if the viral host is not temperate. Interestingly, phages similar to ICP1 (17) have been isolated and shown to infect non-cholera Vibrios (127) which may be candidate viral hosts for the new parasites identified here. Complicating the identification of viral hosts is the arms race that can occur between viral hosts and their parasites, as occurs with ICP1 and PLEs, which may obscure inhibitory activity if the phage being tested has evolved to antagonize the phage parasite in question (15, 102). The outcome of phage infection can be hard to predict based solely on the genome sequences of bacterial hosts and their phages, influencing microbial ecology and complicating phage therapy efforts. Vibrios are responsible for diseases in aquaculture, prompting research into the use of phage therapy to control outbreaks (128). Phage parasites may hinder these efforts, and research into the viral host range and prevalence of putative parasites in aquaculture settings should be considered before application of phages.

Together we have shown that PLE encodes a nuclease that specifically targets ICP1's genome for cleavage, culminating in decreased ICP1 replication and inhibition of ICP1 progeny production. NixI homologs are a feature of an expanded family of putative phage satellites, suggesting that nuclease effectors are a widespread approach used by phage parasites to limit viral host replication. NixI is expressed specifically in response to ICP1 infection, aiding in PLE's efficient cooption of ICP1 components for its own lifecycle. This work furthers our understanding of the tight co-evolutionary relationship between ICP1 and PLE, where adaptations and counter-adaptations between these two genomes have been ongoing for nearly 100 years, shaping phage infection dynamics within cholera-infected patients.

Materials and Methods

Strains and media

A complete list of strains used in this study is available in Supplementary Table 2.1. *V. cholerae* strains used in this study are derivatives of E7946. ICP1_2006_E engineered to lack CRISPR-Cas (Δ CRISPR, Δ cas2-3) was used throughout this study and is referred to ICP1 for simplicity (82). Routine culturing of *V. cholerae* was done in LB at 37°C with aeration. Cells grown in Terrific Broth (TB) were used for protein purification. Antibiotics were supplemented as appropriate at the following concentrations: 50 mg/ml kanamycin for maintenance of the pSUMO vector, 100 mg/ml spectinomycin, 1.25 mg/ml or 2.5 mg/ml chloramphenicol for broth or plate conditions, respectively.

Cloning

Plasmids to express *nixI* and *nixI*^{N95A} are pMMB67EH derivative plasmids with a theophylline inducible riboswitch (Riboswitch E) as described previously (82), and were constructed using Gibson assembly. For expression of *nixI* homologs from *V. cholerae* YBA_A06 (WP_057552372) and *Vibrio parahaemolyticus* S042 (WP_029837635), the genes were commercially synthesized by Genscript and cloned into the pMMB67EH derivative. High copy pUC19 derivative plasmids were used as DNA substrates for nuclease assays. Deletions in

V. cholerae were generated through natural transformation as described previously (83). Naturally competent *V. cholerae* were transformed with PCR products generated by splicing by overlap extension PCR containing arms of homology flanking *frt* recombination sites as described previously (83). All deletions and constructs were verified by Sanger sequencing.

Protein purification

Protein purifications were carried out in Buffer A (150mM Tris-Cl pH=7.8, 50mM NaCl, 2mM BME, 10mM NaCl) unless otherwise indicated or modified. His-SUMO tagged NixI^{N95A} was purified from *Escherichia coli* strain BL21. Catalytically active His-SUMO tagged NixI was purified from *E. coli* BL21 pLysS background to reduce leaky expression of *nixI* which is toxic to *E. coli*. For each *E. coli* expression strain, a 100mL starter culture in TB was grown overnight at 37°C on a shaking incubator at 250 rpm with the appropriate antibiotic(s): kanamycin at 50 mg/ml to maintain the pSUMO plasmid, and chloramphenicol at 25 mg/ml to maintain the pLysS plasmid. The next day, four 1L baffled flasks of TB with appropriate antibiotic(s) were each inoculated with 20mL of overnight starter culture, and grown at 37°C with shaking. When the A₆₀₀ reached 0.5-0.7, the temperature of the incubator was reduced to 18°C, 0.5mM IPTG was added to induce protein expression, and cultures were incubated for 20 hours with shaking. Cells were harvested by centrifugation at 4000xg for 15 minutes then resuspended in 25mL of lysis buffer: 150mM Tris-HCl pH=8, 50mM NaCl, 2mM BME, 1X Pierce phenylmethylsulfonyl fluoride protease inhibitors + 0.5% Triton-X, 1% glycerol. The cell suspension was sonicated for a total of 5 minutes as cycles of 10-second “ON” pulses with 20-seconds “OFF”. After sonication, cell debris was removed by centrifugation at 20,000xg for 40 minutes, the supernatant was filtered using GE 0.2mm regenerated cellulose membranes. Lysate was bound to a GE His-Trap and pumped through using a peristaltic pump. A high salt wash including 2M NaCl was applied to remove DNA. The His-SUMO tagged proteins were eluted in 250mM imidazole in 150mM Tris-Cl pH=7.8, 50mM NaCl, 2mM BME. The eluted protein was then dialyzed against Buffer A + 5% glycerol overnight using SnakeSkin (Thermo Scientific) at 10kDa MWCO dialysis membrane. Dialysis was performed in concert with cleavage of the His-SUMO tag using the SUMO protease (Sigma Aldrich) at 20U/1mg of protein. To remove the SUMO protease and the cleaved tag, the protein was re-applied to Ni-NTA resin and a batch purification was performed, where the cleaved protein of interest remained unbound in the buffer. Protein concentration was measured using a BioPhotometer D30 (Eppendorf). Additional purification of the Ni-affinity purified NixI and NixI^{N95A} was carried out using a Heparin column, equilibrated with Buffer A except with a higher NaCl concentration of 150mM. The column was washed with Buffer A+150mM NaCl until no protein was detected in the flow through, then eluted with 2M NaCl in Buffer A, and dialyzed against Buffer A overnight.

Electrophoretic mobility shift assays (EMSAs)

NixI^{N95A} was incubated with 100bp probes amplified from ICP1, internal to *gp156* and from pGp156* with the putative recognition motif mutated from GTAATCTT to TGTATAGT. Primer sequences are available in Supplementary Table S2. 15nM of DNA probe was added for each 20μL reaction, and incubated with varying concentrations of NixI^{N95A}. Reactions were carried out for 30 minutes at 30°C and then chilled at 4°C. Reactions were run on an 8%

polyacrylamide gel, pre-run at 120V for 60 minutes. EMSA reactions were run at 140V for 40 minutes at 4°C, then visualized by staining with GelRed.

Nuclease assays

Nuclease assays with NixI and NixI^{N95A} were carried out in 150mM Tris-Cl pH=7.8, 50mM NaCl, 1mM DTT and 10mM MgCl₂. Plasmid DNA substrates were added at 100ng per 20μl reaction, with NixI at varying concentrations as indicated. Reactions were incubated at 30°C for 15 minutes, then 1μg/mL of Proteinase K was added to each reaction to digest bound protein and incubated at 37°C for 30 minutes. The entire reaction volume was then run on a 0.8% agarose gel stained with GelRed. To generate cleaved controls, 100ng of vector was digested for 1 hour at 37°C with 10 U of BamHI-HF or EcoRI-HF (NEB) to create a single double-stranded cut or 10 U Nb.BtsI (NEB) to create a single nick, then the enzymes heat inactivated at 80°C for 20 minutes.

Densitometry measurements were performed using ImageJ to quantify nuclease activity on plasmid substrates. Percent cleaved was calculated as the intensity of the nicked+linearized plasmid divided by the total band intensity in each lane. Densitometry measurements reported here are the average of three replicate nuclease assays conducted on different days.

Sample preparation for DNA sequencing and reads mapping

ICP1 infected PLE(+) Δ nixI *V. cholerae* samples were prepared in parallel with previously analyzed PLE(+) and PLE(-) *V. cholerae* samples, as described previously (101). Briefly, 2ml LB cultures of each *V. cholerae* strain was infected at A₆₀₀=0.3 with ICP1 at a multiplicity of infection (MOI) of 1, and aliquots taken at given time points from which total DNA was extracted for sequencing. The pellets were washed with 1mL ice cold phosphate buffered saline, and total DNA was extracted using the QIAGEN DNeasy Blood and Tissue kit. Library prep was done using the NEBNext Ultra II FS DNA Library Prep Kit, and sequenced using Illumina HiSeq4000 at University of California Berkeley QB3 Core Facility. Illumina sequencing reads for each sample and time point were mapped as before (101). Briefly, reads were mapped onto ICP1 reference genome using Bowtie 2 v2.3.4.1 using the default settings except for “-end-to-end” and “-very-sensitive”. Read coverage was normalized by the total number of reads that mapped to the reference.

In vivo cleavage site identification

Illumina reads from samples at 16 minutes post-infection were analyzed for indications of DNA cleavage from PLE(-), PLE(+) and PLE(+) Δ nixI *V. cholerae*. Cleavage ratios in ICP1's genome were calculated following an approach adapted from Culviner et al. (107). To look for putative cleavage sites, the log₂ ratio of reads coverage averaged across three biological replicates at each nucleotide position in ICP1 was calculated for the following comparisons: PLE(+):PLE(-) and PLE(+) Δ nixI:PLE(-). The difference in log₂ coverage was determined for each nucleotide position by subtracting the coverage at the nucleotide before and after each position to determine coverage decrease in both a 5' to 3' and 3' to 5' orientation. These differences were then ranked, identifying 63 putative sites with coverage change >log₂(0.026) (Supplementary Figure 2.1A).

Efficiency of center of infection assays

Cultures of *V. cholerae* were grown in LB supplemented with chloramphenicol (Cm) to maintain the plasmid expressing NixI, NixI^{N95A} or the empty vector control. At an $A_{600}=0.2$, expression from the plasmid was induced with 1mM IPTG + 1.5mM Theophylline. At $A_{600}=0.3$, approximately 15 minutes post-induction, 2mL cultures were infected at MOI=0.1 and incubated for 7 minutes to allow adsorption of ICP1. The infected culture was then diluted 1:2500, 1:25000 and 1:250000 into pre-warmed LB supplemented with Cm and inducer to a final volume of 2ml. Immediately after dilution, 500ml of 1:2500 diluted sample was removed to quantify the input phage. To quantify phage produced, 500ml of each dilution was collected 40 minutes after the initial infection. To all samples, 20ml of chloroform was added to each sample and debris and chloroform was removed by centrifugation at 5,000xg for 15 minutes. Supernatant was used to quantify phage by plaque assay on a permissive PLE(-) host. Efficiency of center of infection was quantified as the ratio between the number of phages produced ml^{-1} minus the number of input phage, from the strains of interest compared to the control culture: either an empty vector control when assessing restriction by *nixI* or wildtype PLE.

Quantification of genome replication by real-time quantitative PCR

ICP1 and PLE replication was quantified by qPCR as described previously, where strains used for quantifying ICP1 replication were infected at an MOI=0.1 and strains used for quantifying PLE replication infected at an MOI=2.5 (32). For strains expressing plasmid constructs, a 2mL culture of interest grown with chloramphenicol was induced with 1mM IPTG 1.5mM theophylline at $A_{600}=0.2$, approximately 15 minutes before infection when $A_{600}=0.3$. A $T=0$ aliquot was immediately boiled to stop DNA replication. Cultures were incubated for 20 minutes at 37°C with aeration, then a 20ml aliquot was boiled. DNA replication was determined by quantifying the fold change in replication at $T=20$ compared to genome copy at $T=0$. To determine the fold change, the C_q value was compared to that of a standard curve of known concentrations of ICP1 or PLE genomic DNA. ICP1's genome was amplified with primers Zac68/69 and PLE replication with primers Zac14/15 using iQ SYBR Green Supermix (Bio-Rad) and the CFX Connect Real-Time PCR Detection system (Bio-Rad). Replication efficiency was compared to a PLE(-) permissive strain harboring an empty vector and similarly grown in the presence of chloramphenicol and inducer.

Identification of an expanded family of PLEs

An HHBlits search with PLE1's NixI protein sequence returned significant homology across the protein sequence to a protein found in non-O1 *V. cholerae* (WP_057552372). The identical protein was found in a total of 10 non-O1 *V. cholerae* isolates isolated in Massachusetts. Analysis of the 10kb regions flanking the NixI homolog showed shared synteny with PLEs, specifically these regions are characterized by a 5' integrase, a total sequence length of ~17kb, and an ORF-less region between divergently transcribed operons. To identify additional putative phage satellites, a PSI-BLAST search excluding *V. cholerae* was performed with NixI from *V. cholerae*'s PLE1 as well as the NixI homolog from non-toxicogenic *V. cholerae* YB2_A06 (WP_057552367.1). We considered proteins potential homologs of NixI if they

shared >25% amino acid identity across >85% of the protein length (Supplementary Table S3). To ensure that annotations were called correctly, the putative phage satellites from *V. parahaemolyticus* s042 and *V. diazotrophicus* 60.18M were verified with RAST.

Statistical and visual analyses

Statistics and graphs were generated using Prism version 9.0.2. Gene maps were made using the gggenes package in R, version 4.0.0. Protein alignments and phylogenetic trees were made using MUSCLE within Jalview version 2.11.0. Consensus motifs of NixI cleavage sites were produced by the WebLogo server (<https://weblogo.berkeley.edu/logo.cgi>).

Acknowledgements

We would like to thank all current and former members of the Seed Lab for their many helpful conversations and feedback. Specifically, we would like to thank Dr. Angus Angermeyer for his help analyzing ICP1's coverage ratios. We would also like to thank the Komeili lab for use of their AKTA for protein purification and the University of California, Berkeley QB3 Core Facility for assistance with whole genome sequencing.

Chapter 4

Toxicity of a PLE-encoded nuclease is modulated by a putative proteolytic adaptor protein

Summary

Prokaryotic viral defense systems deploy nuclease effectors to target and degrade foreign nucleic acids. Of critical importance is the ability for these nucleases to distinguish between invading and host DNA, to prevent self-targeting and death in the absence of infection. Epidemic isolates of *Vibrio cholerae* faces persistent predation from the lytic phage ICP1, and consequently possesses potent defenses against ICP1. PLE is an anti-ICP1 island found in *V. cholerae*, which is activated specifically in response to ICP1 infection. PLE is a parasite of ICP1, and hijacks ICP1 resources for its own lifecycle, resulting in complete inhibition of ICP1 progeny production. One mechanism PLE uses to inhibit ICP1 is a recently discovered nicking endonuclease, NixI, which cleaves ICP1 as the phage attempts to replicate. How PLE is protected from NixI cleavage remained unclear. Although NixI has intrinsic preference for a sequence motif, this motif alone is not sufficient direct NixI to cleave PLE. Here we address this unanswered question by identifying a small co-transcribed protein, StiX, that abrogates NixI toxicity *in vivo*. *In vitro*, StiX did not alter NixI cutting, suggesting it does not prevent NixI DNA binding or cleavage activity. Rather, StiX requires the protease ClpXP to abrogate NixI toxicity, suggesting it may act as an adaptor protein to enhance NixI turnover *in vivo*. StiX is required for wild-type levels of PLE replication and transduction, and defects apparent in the $\Delta stiX$ background are exacerbated when PLE contains sequence motifs that serve as substrates for NixI's nicking activity. We propose a model where StiX contributes to PLE protection from NixI by regulating NixI protein levels or ensuring NixI is degraded at the appropriate time to ensure only ICP1 is targeted.

Introduction

Nuclease effectors are ubiquitous weapons in the arms race between phages and their bacterial hosts (7, 9, 10, 129). Many prokaryotic anti-viral defense systems employ nucleases which specifically recognize and cleave viral DNA (7, 9). In turn, viruses utilize their own suite of nucleases to degrade their hosts' chromosomes and/or to counter host-encoded defenses (118, 124). Indeed, the pressure experienced by phages to evade degradation by nucleases has selected for the evolution of diverse strategies to modify their own genome through nucleotide substitution, as in phage T4 (130), or even to sequester their own genome into an elaborate proteinaceous shell within the host cytoplasm (131). Although nucleases are effective weapons against phages, hosts must protect themselves from self-targeting by constitutively expressed nucleases. Nucleases that possess sequence specificity often detect short DNA motifs that occur frequently, by chance, on host genomes. A common strategy to avoid self-targeting is to differentially modify, often through methylation, of nucleotides at the recognition motifs found on host chromosomes, as is the case for restriction-modification systems (132). As many new defense systems with novel mechanisms of restriction are described in bacterial genomes, it is likely that new ways of distinguishing foreign and self-DNA will be uncovered.

The human pathogen *Vibrio cholerae* is predated upon by the lytic phage ICP1 within cholera-infected human hosts, as both *V. cholerae* and ICP1 have been recovered from patient stool samples. Long-term sampling efforts, dating back over 20 years, demonstrates that *V. cholerae* persists in the face of ongoing ICP1 predation, suggesting cholera encodes defense mechanisms against ICP1 (14). Indeed, *V. cholerae* possesses a subviral parasite known as PLE that limits titers of ICP1 in the population and is a prominent feature of the current pandemic (32). PLE is integrated into *V. cholerae*'s chromosome, until the host is infected by ICP1. PLE excises and replicates by hijacking ICP1 resources for these essential processes, and late in infection, PLE packages its genome in modified ICP1 virions to disseminate its genome through transduction (35, 101, 118). Importantly, PLE(+) *V. cholerae* hosts still lyse when infected by ICP1, but instead of releasing ICP1 progeny, infected cells exclusively release PLE transducing particles (32, 100). The mechanisms PLE deploys to inhibit ICP1 are under investigation, though it is appreciated that PLE encodes multiple inhibitory mechanisms (32, 35).

One recently uncovered PLE-encoded anti-phage mechanism is the nicking endonuclease, NixI (89), which introduces nicks into ICP1's genome concurrent with ICP1's attempted transition to rolling-circle replication (RCR): an essential process in the phage lifecycle required to linearize its genome into concatemers for packaging (121). We previously proposed that NixI introduces nicks into ICP1 as it transitions to RCR, breaking DNA concatemer formation leading to a failure to package its genome (89). However, PLE also undergoes a transition to RCR to linearize its genome for packaging. How PLE is protected from NixI-mediated cleavage was unclear. We anticipated that PLE may be protected in part through NixI's sequence preference for the motif GNAANC[^]TT (where cleavage occurs at the arrow), which is absent from PLE1 (used in this study) and found 99 times in ICP1's genome (89). However, PLEs 2 and 10, which encode amino acid identical copies of PLE1's NixI, possess three and four instances of this motif, respectively, yet replicate and transduce efficiently (32, 33). Taken with our finding that PLE1 engineered to contain NixI's preferred cleavage motifs also showed no defect in replication or transduction, we postulated that additional mechanisms were likely required to ensure PLE remains protected from NixI activity for efficient mobilization. Here, we look for additional PLE-encoded factors that may contribute to protection from NixI cleavage. This led us

to identify StiX, which abrogates NixI toxicity *in vivo*. We report here that StiX is required for PLE replication and transduction, particularly when PLE contains NixI recognition motifs. We propose a mechanism of StiX function wherein StiX acts as an adaptor to bring NixI to the protease ClpXP for degradation, contributing to PLE's protection from NixI.

Results

PLE's nuclease effector is neutralized by the small co-expressed protein StiX

Upon infection by ICP1, PLE undergoes a temporally regulated transcriptional program (59), typical of lytic and temperate phages. To gain insight into how PLE is protected from NixI nuclease activity, we analyzed previously published RNA-seq analyses of ICP1 infecting PLE(+) *V. cholerae* to identify co-transcribed PLE-encoded open reading frames (ORFs) as candidate genes that may alter NixI function. We identified a small downstream ORF of interest that is co-transcribed with NixI, herein referred to as *stiX*. In the absence of infection, transcription of *nixI* and *stiX* is tightly repressed, consistent with NixI's previously identified role as an endonuclease, which could compromise cell viability if expressed even at low levels. Transcription of *nixI+stiX* initiates around 8 minutes post infection (mpi), and peaks between 12- to 16-mpi by ICP1 (59) (Figure 4.1A), showing that *stiX* is co-expressed with *nixI*. To determine whether StiX functions in concert with NixI, the genes were cloned under the same inducible promoter, retaining the native orientation of *nixI* closer to the promoter. Strikingly, co-induction of *nixI+stiX* rescued colony formation of *V. cholerae* though colonies were smaller compared to the empty vector control (Figure 4.1B). To quantify *stiX*-mediated rescue, strains expressing either *nixI*, *nixI+stiX* or an empty vector were induced, and cell viability was measured 30 minutes post-induction. We observed that cell viability was restored when *stiX* was co-expressed with *nixI* (Figure 4.1C), suggesting that StiX may abolish or restrict NixI nuclease activity *in vivo*.

Like many PLE-encoded ORFs, we were unable to detect any conserved domains nor homologs of StiX outside of PLEs using BLAST. Structural predictions of StiX generated by AlphaFold (133) yielded a high confidence model of StiX as a penta-helical bundle helix-turn-helix domain-containing protein which is archetypal of prokaryotic transcription factors (134) (Figure 4.2A). Using HHpred (135), we found that StiX shares predicted structural homology with the C-terminal domain of the alternative extracytoplasmic sigma factor (ECF) SigE (Probability score=98.06%, e-value<0.0003, 21% total amino acid identity). Specifically, predicted structural homology was detected between StiX and the σ^4 domain of SigE, the domain from an *E. coli* ECF that contacts the -35 promoter element but does not participate in strand separation (136). These structural predictions point to a role for StiX in DNA-binding, prompting us to test whether StiX binds DNA. Given that StiX abrogates NixI toxicity *in vivo*, we purified StiX to test whether it can bind DNA, hypothesizing that StiX binding may occlude NixI from its target motifs. Contrary to our prediction, StiX did not bind to DNA probes from regions of ICP1's genome (in *gp121* or *gp156*) which were previously shown to be bound by NixI under the same conditions (Figure 4.2B). Given that StiX shares functional similarity to antitoxins, we next tested whether StiX binds its own promoter, as many antitoxins regulate transcription of the toxin-antitoxin module (137). Of note, the toxicity assays performed here relied on inducible expression of *nixI+stiX* from a non-native promoter, allowing us to rule out the possibility that StiX abrogation of NixI's toxicity *in vivo* was not due to repression of *nixI* expression. Considering the shared role of antitoxins in auto-regulation and StiX's homology to ECF sigma

factors, we predicted that StiX may bind to its own promoter. However, we were unable to detect any DNA binding of StiX to a DNA probe spanning the *nixI* promoter region which we had identified by RNA-seq (Figure 4.2C). Given initial concerns over the purity of our first affinity purification of StiX, we re-purified StiX and tested the ability of both protein preparations to bind DNA (Supplementary Figure 3.1). With both purifications, we did not detect any DNA binding by StiX. Although we cannot rule out that StiX may bind other sequences of DNA, these data suggest that StiX does not abrogate NixI's toxicity by binding and occluding NixI from its target motifs.

We then hypothesized that StiX may directly bind to NixI and prevent it from cleaving DNA targets. We tested the ability of purified NixI to cleave plasmid targets *in vitro* in the presence and absence of StiX. Strikingly, we found that even with an excess of StiX, NixI nicked both an empty vector substrate (Figure 4.3A) as well as a plasmid containing NixI target sequence (Figure 4.3B) in the presence of StiX compared to without StiX. This suggests that StiX is not sufficient to inhibit NixI nuclease activity *in vitro*, showing that StiX neither prevents off-target cleavage of an empty vector substrate by NixI, nor does it prevent cleavage of its preferred substrate.

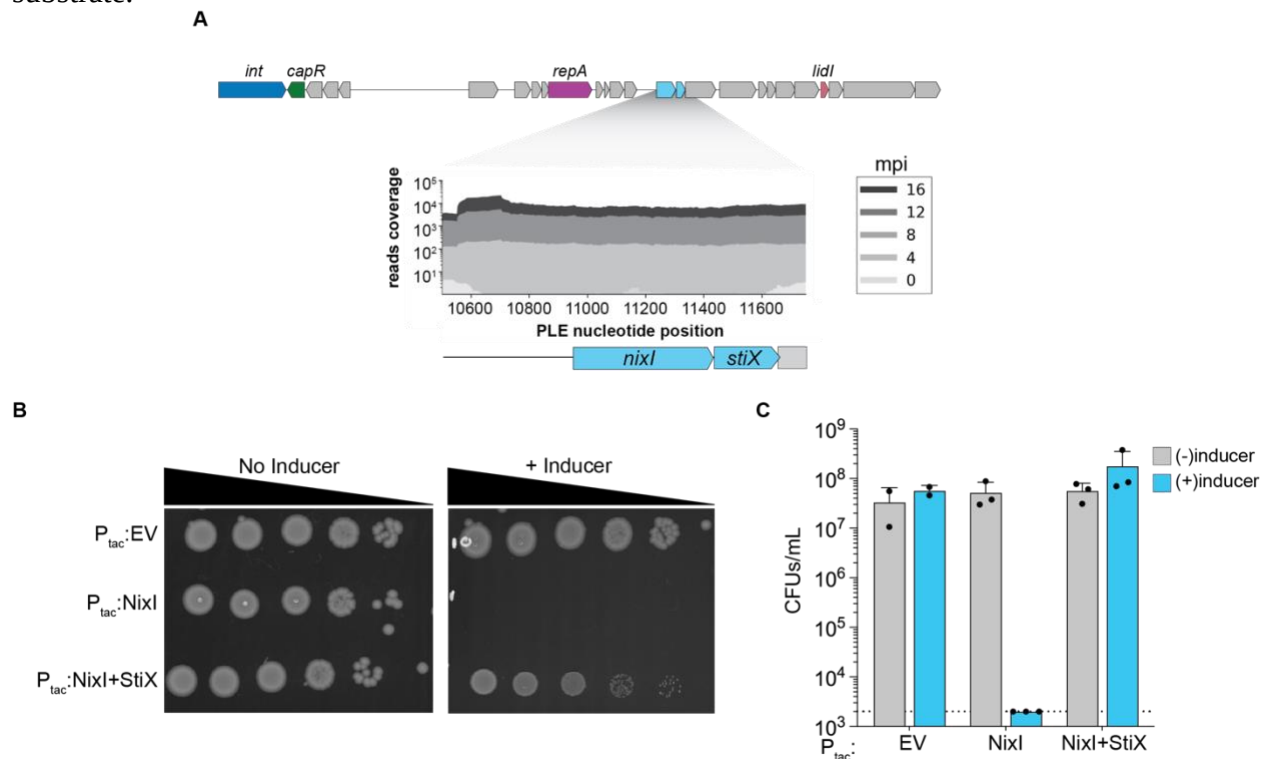


Figure 4.1. NixI is co-transcribed with a small downstream open reading frame, StiX, that abrogates its toxicity. (A) Top: gene map of PLE 1 where genes with previously described functions are colored. Bottom: Previously published (59) RNA-seq reads of PLE1 transcripts mapped onto the inset region containing the promoter for the *nixI*+*stiX* operon. Reads are colored by time post-infection according to the legend (mpi=minutes post-infection). Reads accumulate over the course of infection, showing reads mapping to *nixI*+*stiX* increase substantially between 8 and 12 mpi. (B) 10-fold dilution series of *V. cholerae* with inducible plasmid constructs expressing *nixI* or *nixI*+*stiX* spotted onto agar plates without and with inducer. (C) Quantification of cell viability of *V. cholerae* expressing *nixI* or *nixI*+*stiX* plated at the time of induction (grey bars) and 30-minutes post-induction (blue) onto plates without inducer. Bar height is mean, whiskers are standard deviation and dots represent biological replicates.

StiX can only abrogate NixI nuclease activity in the presence of the protease ClpXP

To reconcile our findings that StiX abrogates NixI toxicity *in vivo*, with *in vitro* data showing no effect of StiX on NixI-mediated nuclease activity, we considered whether an additional protein partner may alter NixI toxicity in the presence of StiX. In addition to their canonical roles as DNA-binding proteins, HTH domains have been shown to mediate protein-protein interactions (134, 138). Given that StiX shows functional similarity to antitoxin proteins, we were intrigued by the mechanism of the *Caulobacter crescentus* encoded SocA antitoxin that abrogates SocB toxicity by serving as an adaptor to mediate ClpXP proteolysis (139). SocA recruits ClpXP to the toxin SocB, leading to its degradation (139). In the absence of SocA, SocB is more stable and causes an arrest in cell growth by inhibiting DNA replication (139). We hypothesized that StiX may act as a ClpXP adaptor, mediating NixI degradation. To test for a role for ClpXP, we assayed the toxicity of plasmid constructs expressing either *nixI* or *nixI+stiX* in *V. cholerae* $\Delta clpP$ to determine whether StiX could abrogate NixI toxicity in the absence of functional ClpXP. Strikingly, StiX failed to abrogate NixI toxicity in the absence of ClpXP, phenocopying the toxicity of NixI alone in a *V. cholerae* $\Delta clpP$ host (Figure 4.4A). This suggests that StiX abrogates NixI toxicity in a ClpXP dependent manner, perhaps by acting as an adaptor for NixI proteolysis.

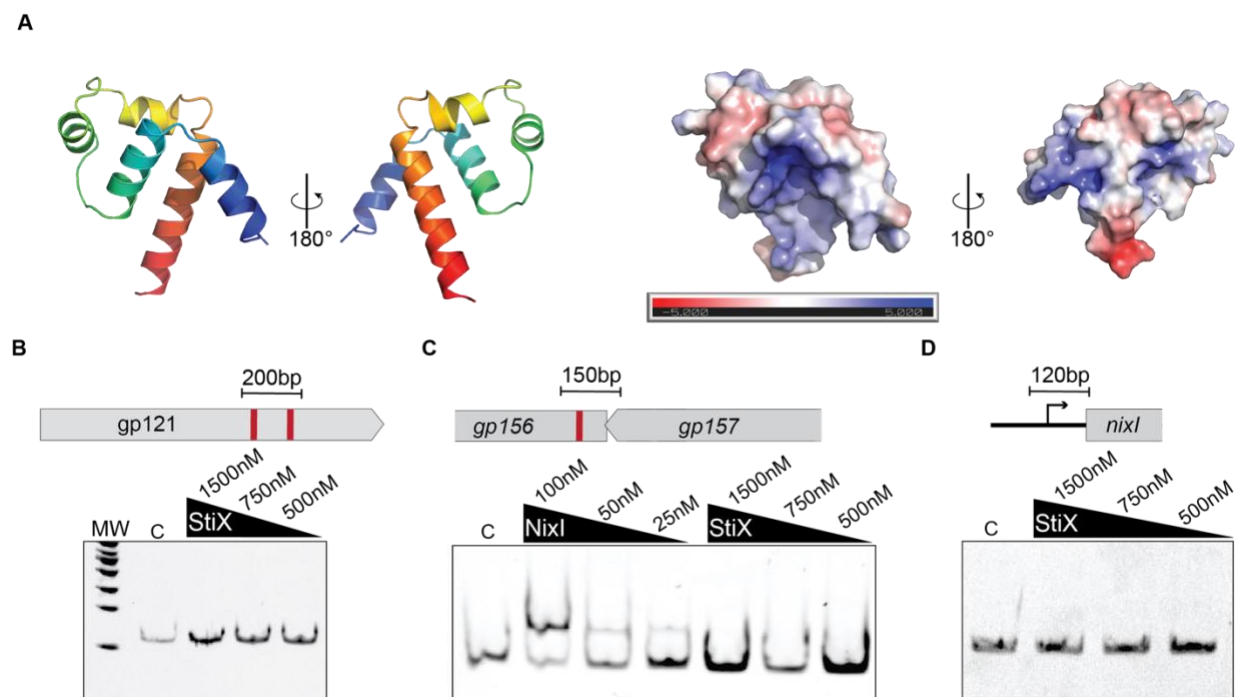


Figure 4.2. StiX does not demonstrate DNA-binding activity *in vitro* despite containing predicted DNA-binding domains. (A) Left: Predicted structure of StiX generated by AlphaFold colored with the N-terminus as blue and the C-terminus in red, shown rotated around the vertical axis. Right: Electrostatics of the predicted structure for StiX shows a small positively charged pocket suggesting a potential interaction surface for DNA. (B) DNA binding assayed through an electrophoretic mobility shift assay (EMSA) of StiX for a probe spanning NixI's cut sites in ICPI's *gp121*, where the position of NixI's cleavage motif is colored in red. The concentration of StiX is shown about the arrow on top of the gel. C=probe without added protein, MW=molecular weight ladder. (C) as in B, but with a probe spanning NixI's cut sites in ICPI's gene *gp156*. NixI binding to this probe is shown for comparison as previously described (89). (D) as in B, except displaying StiX binding of a DNA probe spanning NixI's promoter.

StiX contributes to protecting PLE from NixI cleavage

The initial description discovery of NixI as an ICP1-targeting endonuclease failed to explain how PLE evades NixI cutting itself. A component of PLE protection may come from nucleotide sequence, as PLE1 does not contain the consensus motif required for optimal NixI nuclease activity. However, NixI nuclease activity is promiscuous, and will cleave non-target sequence at high concentrations *in vitro* (89). Additionally, PLE1 was protected from NixI (as assessed by PLE1 replication and transduction) even when engineered to contain motifs from ICP1 shown to be robustly cleaved by NixI *in vivo* and *in vitro* (schematic shown in Figure 4.4B). This engineered PLE, termed PLE+*gp121*, was able to replicate and transduce to wildtype levels. We hypothesized that StiX may ensure rapid turnover of NixI through ClpXP-mediated proteolysis to protect PLE from self-targeting, giving NixI only enough time to cleave its preferred sites in ICP1. To test this, we asked whether StiX is required for PLE replication in a wildtype PLE1 background as well as in our engineered PLE1 containing the region from *gp121* in ICP1, which is a primary target of NixI (89). PLE1 Δ *stiX* showed only a modest replication defect, suggesting it is not part of PLE's replisome (Figure 4.4C), which we anticipate would show a much larger replication defect as seen when PLE's replication initiator protein RepA is knocked out (101). However, a replication defect of the magnitude seen in a Δ *stiX* background suggests StiX may function to protect PLE from NixI cleavage. This is because the qPCR primers used to detect PLE replication amplify only a 150bp region of PLE, meaning this assay does not assess genome integrity. In support of this hypothesis, PLE1+*gp121* Δ *stiX* showed an even greater replication defect, consistent with NixI cutting of the *gp121* motifs in the absence of StiX (Figure 4.4C).

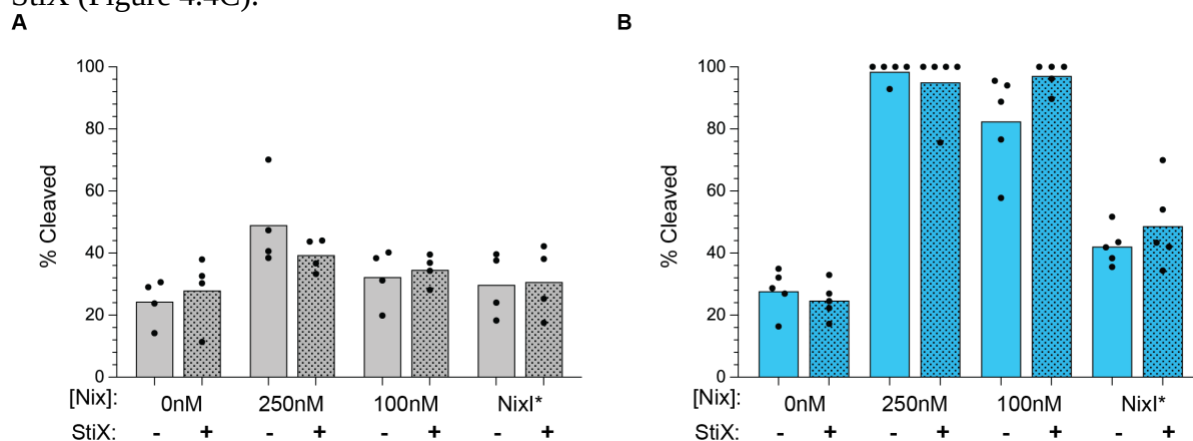


Figure 4.3. StiX does not inhibit NixI nuclease activity *in vitro*. (A) Densitometry measurements calculating the percent of empty vector plasmid cleaved out of total band intensity per lane from replicate *in vitro* nuclease assays. The concentration of NixI is displayed on the x-axis. StiX was added at 500nM and a purified catalytically dead NixI* at 500nM. (B) As in B, except displaying the percent of plasmid containing a region from *gp121* shown previously to be robustly cleaved by NixI *in vitro* and *in vivo*. For A and B, bar height is the mean and dots represent independent nucleic acid assays.

We next asked whether the replication defect observed when PLE lacks *stiX* would manifest as a reduction in PLE's ability to package its genome and mobilize through horizontal transduction. NixI targeting of PLE in a PLE Δ *stiX* background may result in a transduction defect greater than the replication defect seen by qPCR, as genome integrity is essential for proper packaging and horizontal spread. When normalized to wildtype PLE1 transduction, PLE1 Δ *stiX* exhibited a substantial transduction defect of nearly an order-of-magnitude, which could be

complemented by expression of *stiX* *in trans* (Figure 4.4D). PLE1+*gp121* Δ *stiX* showed a similar transduction defect as PLE Δ *stiX*, which could also be complemented by expression of *stiX* *in trans* (Figure 4.4D). Together these data suggest that in the absence of StiX, NixI may target PLE, which can be exacerbated when PLE was engineered to contain NixI's preferred target motifs.

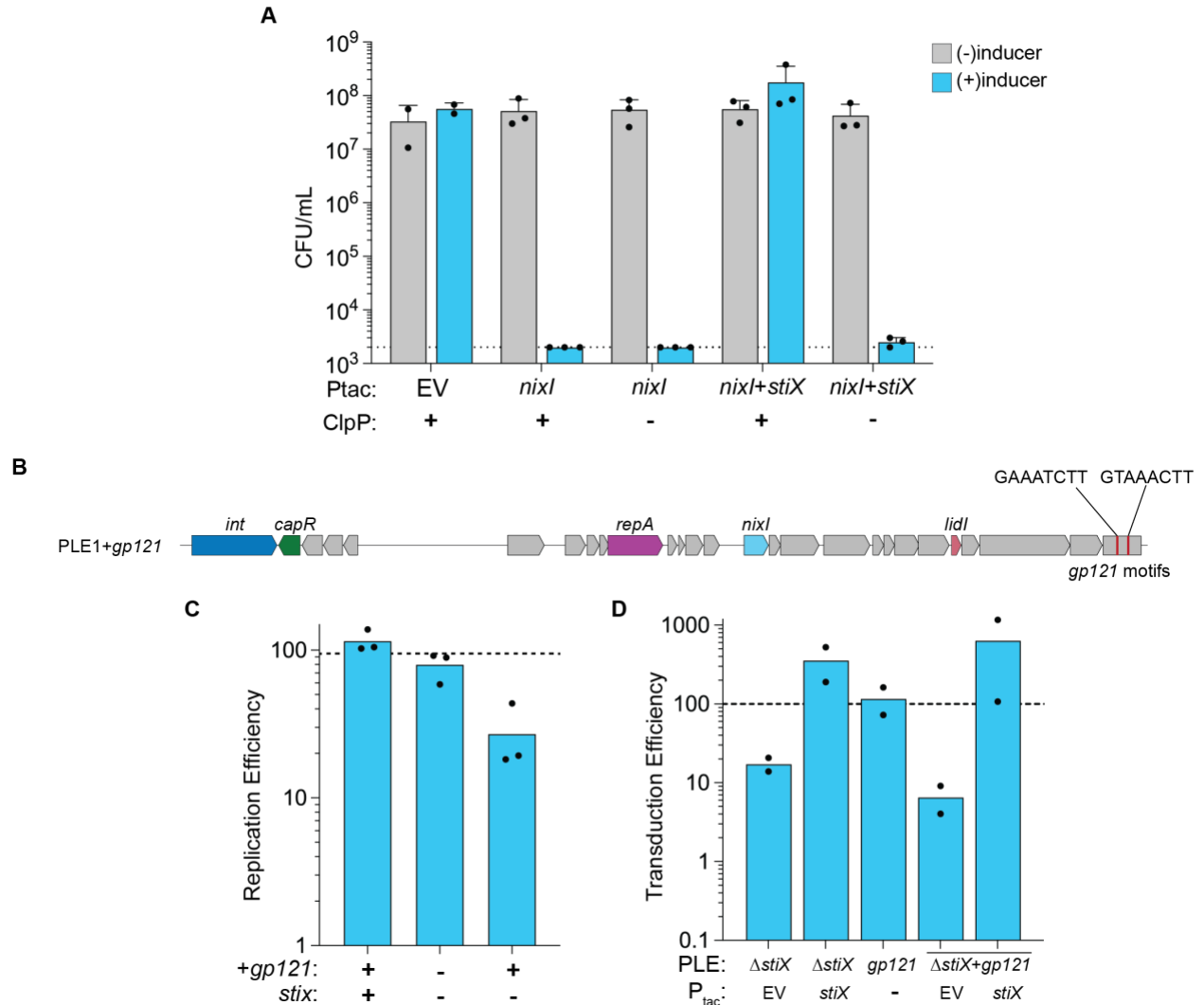


Figure 4.4. StiX may protect PLE from NixI cleavage by promoting ClpXP-mediated proteolysis. (A) Cell viability of *V. cholerae* expressing *nixI* or *nixI*+*stiX* in a wildtype or Δ *clpP* background plated at the time of induction (grey bars) and 30-minutes post-induction (blue). Wildtype data replicates are the same as shown in Figure 1C, which were done at the same time as the Δ *clpP* samples shown here. Bar height is mean, whiskers are standard deviation and dots represent biological replicates. (B) Gene map of PLE1+*gp121*, showing the location at the end of PLE1 where NixI's target region from ICP1's *gp121* was cloned (89). (C) Replication efficiency of PLE with *gp121* or where *stiX* was deleted displayed as the percentage of wildtype PLE replication (dashed line). (D) Transduction efficiency of PLEs with the *gp121* region or with *stiX* deleted, where *stiX* was complemented *in trans* from a plasmid, shown as a percent of wildtype PLE transduction (dashed line). For C and D, bar height is mean and dots display independent biological replicates.

We propose a model (Figure 4.5) where upon ICP1 infection of a PLE(+) *V. cholerae* host, PLE undergoes a temporally controlled transcriptional program to co-express NixI with StiX. NixI cleaves its preferred sites in ICP1 while StiX targets NixI for proteolysis by the

unfoldase/protease ClpXP, perhaps by acting as an adaptor for ClpX. By limiting the half-life of NixI, StiX helps to protect PLE from self-cleavage, ensuring wildtype PLE replication and transduction.

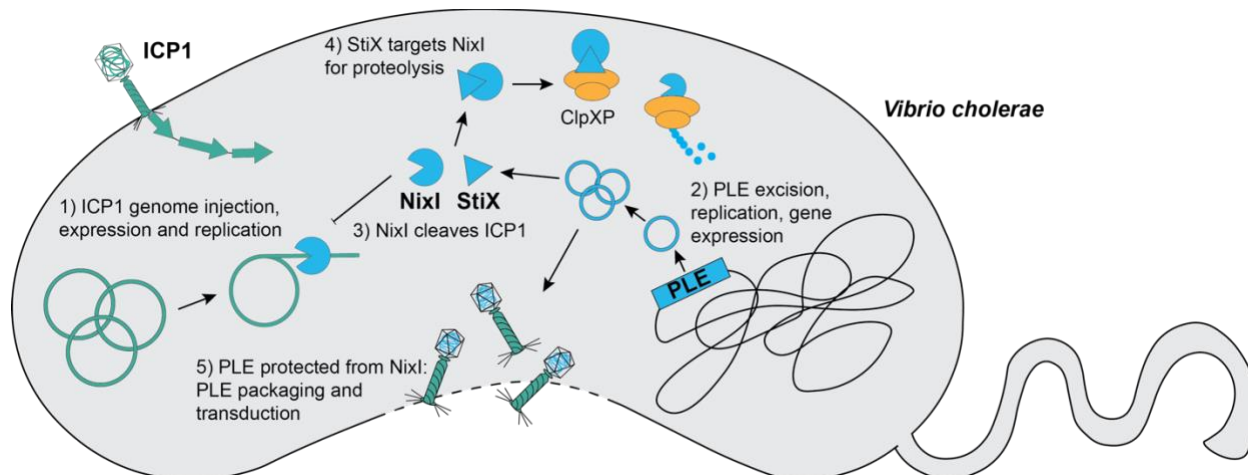


Figure 4.5. Proposed model of how StiX protects PLE from NixI mediated degradation. Upon infection by ICP1, PLE excises from *V. cholerae*'s genome and undergoes a temporally regulated transcriptional program, concurrent with ICP1 replication and gene expression. PLE expresses NixI which cleaves ICP1's genome at the time when ICP1 is attempting to transition to rolling-circle replication. StiX is co-transcribed with NixI and may act as an ClpX adaptor to bring NixI to ClpP for proteolysis. Controlling NixI protein levels may help protect PLE from cleavage by NixI, allowing for proper PLE genome packaging and transduction.

Discussion

Bacteriophages and their parasites, such as ICP1 and PLE, complete rapid lifecycles within their bacterial hosts, culminating in host lysis within remarkably reproducible timeframes (17, 140). Transcriptional profiling using techniques like RNA-seq contributed to an understanding of viral lifecycles as tightly regulated transcriptional programs, which are often simplified into three periods of gene expression: early, middle and late gene expression (59, 141). Early phage genes are often involved in host takeover, while middle genes include functions related to replication and late genes encode for structural components and packaging (141). Successful transition through a phage's lifecycle requires both temporal gene expression but also turnover of proteins that have completed their functions and are no longer needed. To our knowledge, mechanisms that regulate proteolysis of phage proteins across a temporal gradient are under-explored. Here we identify a protein, StiX, encoded by the phage parasite PLE, that depends on the protease complex ClpXP for its function. StiX abrogates the toxicity of the PLE-encoded nickase NixI when ectopically expressed, and is required for PLE replication and transduction, particularly when PLE was engineered to encode NixI's preferred recognition motifs. This implicates StiX in turnover of NixI, which may help protect PLE from NixI-mediated cleavage. It is possible that StiX targets additional PLE-encoded proteins for proteolysis and could be a subject of future study. What remains to be determined is how NixI can target ICP1's genome before StiX mediates NixI proteolysis. We propose two possible mechanisms of control, the first through protein levels and the second through tight temporal regulation. First, StiX may control NixI protein levels, ensuring just the right amount remains in

the cell to direct NixI only to its preferred sites in ICP1. Secondly, PLE may initiate RCR slightly later than ICP1, so that only ICP1 is susceptible to replication collapse when nicked by NixI. StiX could then target NixI for degradation before PLE's transition to RCR, ensuring that even if NixI nicks PLE, it would not lead to replication collapse.

At first blush, NixI and StiX may be considered a toxin-antitoxin system. Toxin-antitoxin (TA) systems are small modules composed of a proteinaceous toxin and an antitoxin that can be either a protein or an RNA (142). These small modules are highly diverse and defined by the functional property that the toxin impairs bacterial host growth/viability while the antitoxin abrogates the effect of the toxin (142, 143). For type II TA systems, where both the toxin and antitoxin are proteins, the model of TA function is that of a constitutively expressed TA system, where the antitoxin is highly labile, such that any perturbation to bacterial transcription or translation would lead to rapid degradation of the antitoxin, liberating the toxin to carry out its effects on processes such as inhibiting protein synthesis or cleaving nucleic acids (144). The biological role of toxins is to intoxicate the cell, and this intoxication has been implicated in post-segregation killing (137) and triggering persister cell formation (145). Although ectopic expression of NixI leads to cell death, *nixI* is only expressed during infection by ICP1 in a context where NixI's ability to intoxicate its host must be limited, as PLE1's lifecycle depends on robust replication (101) and transcription (59) and is unaffected by NixI. Many TA systems have now been implicated in phage defense (146), and NixI cleavage of the phage ICP1 would fit within this paradigm.

C. crescentus encodes the unique type II TA system SocAB, where the antitoxin SocA brings the toxin SocB to ClpXP to enhance toxin proteolysis (139). When the antitoxin is not expressed, the toxic protein SocB impairs DNA replication. This finding prompted us to test here whether StiX may mediate NixI proteolysis in a ClpP-dependent manner. Indeed, we found that StiX requires protease ClpP to abrogate NixI toxicity, suggesting that StiX may function as an adaptor for the unfoldase ClpX. Of note, this discovery was made in the context of ectopic expression of NixI and StiX, and the role of ClpXP in NixI turnover during native expression during ICP1 infection has not been directly tested. All proteins with the potential to cause deleterious effects on their hosts should be tightly and specifically regulated. For many nucleases or genes with a role in temporally controlled processes, cells may express proteins or RNAs that control their mode of action or shorten their half-life. For example, proteins involved in temporally controlled cellular processes, such as cell division, can have deleterious effects on their hosts if they remain in their host cells after they have carried out their function. In *Caulobacter crescentus*, the response regulator CtrA is essential for transcription at some time points in the cell cycle but will inhibit DNA replication if maintained within the cell (147). Proteolysis of CtrA is controlled by ClpXP and accelerated by an adaptor complex that contributes to the rapid turnover of CtrA *in vivo* (147). Yet because the mechanism of CtrA function is well-understood, CtrA with its adaptor are not considered a TA system. Many toxins have yet unknown functions for their host cells, and their identification as TA systems often relies on ectopic expression without knowledge of when and to what levels they would be expressed naturally. TA systems may have more subtle phenotypes when expressed in their native contexts, begging for caution in concluding that the purpose of these proteins is to intoxicate the cell.

Further work is needed to understand the temporal dynamics of NixI turnover. Specifically, quantitative Western blots should be carried out to determine whether NixI protein levels decline in cells expressing StiX, and whether they are restored in a $\Delta clpP$ host background. Additionally,

it remains unknown how NixI can target ICP1 with co-expressed StiX, though we hypothesize that NixI levels may be titrated by StiX so that enough protein is available to cleave ICP1, yet excess protein is turned over by StiX. In our previous work on NixI, we postulated that ICP1 would be particularly vulnerable to nicking by NixI when it is transitioning to rolling-circle replication (RCR). StiX may degrade NixI before PLE attempts to transition to RCR, ensuring that PLE is less vulnerable to NixI nicking when it undergoes its shift in replication strategy. While this likely would contribute to PLE-protection from NixI, there are likely other factors that protect PLE from NixI activity, such as the contribution of sequence motifs and topology of the DNA substrate, which we identified previously (89).

Materials and Methods

Strains and media

A complete list of strains used in this study is available in Supplementary Table 3.1. *V. cholerae* strains used in this study are derivatives of E7946. ICP1_2006_E engineered to lack CRISPR-Cas (Δ CRISPR, Δ cas2-3) was used throughout this study and is referred to ICP1 for simplicity (82). Routine culturing of *V. cholerae* was done in LB at 37°C with aeration. Cells grown in Terrific Broth (TB) (12g Tryptone, 24g Yeast Extract, 4mL of glycerol in 900ml, then 100ml of 10X phosphate buffer was added) were used for protein purification. Toxicity was assessed by inducing expression of *nixI* or *nixI+stiX* from an inducible plasmid (see “Cloning” below) in 2mL liquid cultures for 30 minutes. Colony forming units (CFUs) were quantified on $\text{Sm}^{100}+\text{Cm}^{2.5}$ agar plates prior to induction (at T=0) and 30 minutes post-induction (T=30).

Cloning

A full list of primers used in this study is presented in Supplementary Table 3.2. Induction of plasmids were controlled by P_{tac} and a theophylline inducible riboswitch (Riboswitch E) though labelled here as “ P_{tac} ” for simplicity (82). Plasmids to express *nixI* and *stiX* are pMMB67EH derivative plasmids as described previously (82), and were constructed using Gibson assembly. High copy pUC19 derivative plasmids were used as DNA substrates for nuclease assays. For StiX purification, *stiX* was cloned onto either a His-pSUMO vector or pCold II vector, constructed by Gibson assembly. Deletions in *V. cholerae* were generated through natural transformation of homologous constructs as described previously (83). All deletions and constructs were verified by Sanger sequencing.

Protein purification

Protein purifications were carried out in Buffer A (50mM Tris-Cl pH=8, 200mM NaCl, 5% glycerol) unless otherwise indicated or modified. Assaying StiX activity was done with two independent protein purifications, and protein purifications are shown in Supplementary Figure 3.1. StiX was purified both from a His-SUMO vector with a cleavable N-terminus 6His-SUMO tag and from a pCold vector with a CTD 6His tag. Both proteins were purified from *Escherichia coli* strain BL21. For each *E. coli* expression strain, a 100mL starter culture in Terrific Broth was grown overnight supplemented with the appropriate antibiotic: Kanamycin at 50mg/ml to maintain the pSUMO plasmid, and carbenicillin at 50 mg/ml to maintain the pCold, grown at

37°C on a shaking incubator at 250rpm. The next day, 1L baffled flasks of Terrific Broth were each inoculated with 20mL of overnight starter culture, with appropriate antibiotics and grown at 37°C with shaking. When the OD=0.5-0.7, the temperature of the incubator was reduced to 16°C, and 1mM IPTG was added to induce protein expression. For pCold induction, the culture flasks were cold-shocked in an ice water bath for 30 minutes prior to transfer to the 16°C shaking incubator. Cultures were induced for 20 hours then harvested by spinning at 4000xg for 15 minutes, in a Sorvall Lynx 400 Centrifuge (Thermo Scientific) then resuspended in lysis buffer: 50mM Tris-HCl pH=8, 200mM NaCl, 5% glycerol, 1X Pierce PMSF protease inhibitors + 0.5% Triton-X, 1% glycerol. The cell suspension was sonicated for a total of 10 minutes as cycles of 10-second “ON” pulses with 20-seconds “OFF”. After sonication, cell debris was removed by centrifugation at 26,000 xg for 45 minutes. Lysate was bound to Ni-NTA Sepharose resin. Proteins were eluted in increasing concentrations of imidazole. The eluted protein was then dialyzed against Buffer A + 10% glycerol overnight using a 3kDa MWCO dialysis membrane (Thermo Scientific). Dialysis was performed in concert with cleavage of the His-SUMO tag using the SUMO protease (Sigma Aldrich) at 20U/1mg of protein. To remove the SUMO protease and the cleaved tag, the protein was re-applied to Ni-NTA resin and a batch purification was performed, the cleaved protein of interest remained unbound in the buffer. Protein concentration was measured using a BioPhotometer D30 (Eppendorf). Additional purification of StiX was performed using size exclusion chromatography on the GE AKTA, with fractions containing StiX pooled and concentrated using a 3kDa MWCO centrifugal filter.

Electrophoretic mobility shift assays (EMSAs)

StiX or NixI^{N95A} were incubated with DNA probes amplified from *gp121*, *gp156* or 120bp upstream of the NixI promoter region. 20nM of DNA probe was added for each 20μL reaction, and incubated with varying concentrations purified proteins. Reactions were carried out for 30 minutes at 30°C and then chilled at 4°C. Reactions were run on an 8% polyacrylamide gel in 0.5X TB buffer, pre-run at 120V for 30 minutes. EMSA reactions were run at 140V for 40 minutes at 4°C, then visualized by staining with GelRed.

Nuclease assays

Nuclease assays with NixI and NixI^{N95A} were carried out in 150mM Tris-Cl pH=7.8, 50mM NaCl, 1mM DTT and 10mM MgCl₂. Reactions containing StiX were performed with a constant amount of StiX at 500nM. Plasmid DNA substrates were added at 100ng per 20μl reaction, with NixI at varying concentrations (10-250nM). Reactions were incubated at 30°C for 15 minutes, then 1μg/mL of Proteinase K was added to each reaction to digest bound protein and incubated at 37°C for 30 minutes. The entire reaction volume was then run on a 0.8% agarose gel stained with GelRed.

Densitometry measurements were performed using ImageJ to quantify nuclease activity on plasmid substrates. Percent cleaved was calculated as the intensity of the nicked+linearized plasmid divided by the total band intensity in each lane, and averaged as three technical measurements for each lane. Densitometry measurements were performed on three replicate nuclease assays conducted on different days.

Quantification of genome replication by real-time quantitative PCR

ICP1 replication was quantified by qPCR as described previously (32) in *V. cholerae* hosts that were either PLE1(+), PLE1+*gp121*, or PLE1(+) Δ *stiX*. *V. cholerae* host expressing either an empty vector or *stiX*. For strains expressing plasmid constructs, a 2mL culture of interest grown with 1.25 μ g/ml Cm was pre-induced with 1mM IPTG and 1.5mM theophylline, 20 minutes prior to infection. At OD₆₀₀=0.3, cultures were infected at MOI=0.1, and a T=0 aliquot was immediately boiled to stop DNA replication. Cultures were incubated for 20 minutes at 37°C with aeration, then a 20ml aliquot was boiled to quantify final DNA replication. To quantify fold-change in PLE genome copy, primers Zac14/15 were used with iQ SYBR Green Supermix (Bio-Rad) using CFX Connect Real-Time PCR Detection system (Bio-Rad). Replication efficiency was compared to a wildtype PLE(+) strain.

Transductions

PLE transductions were carried out as described previously (35). Briefly, a *V. cholerae* PLE-Kan^R donor cell was infected when OD=0.3 by ICP1 at MOI=2.5, and incubated at 37°C for 5 minutes to allow for phage attachment. After 5 minutes, unbound phage were washed out and cells resuspended in fresh pre-warmed LB with 10mM of Mg₂SO₄, supplemented with inducer in cases where a knockout was being complemented. Infected cultures were incubated for 30 minutes then chloroform treated with 20 μ l of chloroform and set at room temperature for 10 minutes prior to centrifugation at 4°C at 5,000 $\times$ g to remove cell debris and chloroform. Spectinomycin-resistant recipient cells (Δ *lacZ*::Spec^R) were grown overnight in LB and supplemented with 10mM of Mg₂SO₄, when removed from the incubator ~16 hours after inoculation. Donor lysates were added to recipient cells at a 1:10 and a 1:1 ratio and incubated at 37°C with aeration for 20 minutes. Dilutions of recipient cells were plated and quantified on Kan+Spec plates to select for successful transduction recipients. Transductions were performed in biological triplicates.

Statistical and graphical analyses

Graphs were generated using Prism version 9.0.2. Gene maps were made using the gggenes package in R, version 4.0.0

Acknowledgments

We would like to thank members of the Seed Lab for their insights into this project. Particularly, we would like to thank Zoe Netter, Peter DePaola and Kaitlyn Wang for their valuable contributions to this project. We would also like to thank the Komeili Lab for use of their AKTA for protein purification.

Chapter 5

The future of phage biology: Centering phages in bacterial ecology and evolution

Ongoing surveillance of *V. cholerae* isolates and phages shed from cholera infected patients in endemic regions has led to insights into how phages can shape the ecology and evolution of their hosts, highlighting the prominent role of mobile genetic element (MGE) gain and loss in driving the evolution and diversity of bacterial populations. In this work, we have highlighted how two different families of MGEs contribute to phage resistance in clinical isolates of *V. cholerae*. The first family of MGEs are SXT ICEs: self-transmissible elements that are integrated into their host chromosomes until DNA damage triggers their excision and transfer to a new host. While SXT ICEs have been appreciated for quite some time for their role in conferring antibiotic resistances to *V. cholerae* and other pathogens such as *Proteus mirabilis* (148, 149), we showed here that they also confer phage resistances in one of their cargo gene islands, termed hotspot 5. We found that all SXT ICE genomes available to date contain bioinformatically predictable phage defense systems within hotspot 5, suggesting that phage defense is broadly conferred by SXT ICEs in diverse taxa of Gammaproteobacteria. A given bacterial taxon can possess several different flavors of SXT ICEs within a population. This suggests several different evolutionary outcomes are possible when such a bacterial population encounters a changing biotic or abiotic environment.

First, certain SXT ICE cargo genes may confer a fitness advantage in particular environments, selecting for one SXT ICE over another. This could lead to a selective sweep, where a bacterial strain with a given SXT ICE can bloom and become the dominant lineage in a new environment. For example, here we showed that changes to a bacterium's biotic environment, as in when phages adapt to overcome hotspot 5 encoded defenses of a particular SXT ICE may contribute to its replacement by an SXT ICE to which the phage are naïve and therefore inhibited. The same may prove to be true for changes to abiotic environmental conditions, such as the application of antibiotics in populations where SXT ICEs confer different antibiotic resistance profiles. Treatment of a bacterial infection may select for one SXT ICE over another and lead to a subsequent lineage shift. Where there is other diversity between strains or within other SXT ICE cargo islands, these sweeps may bring with it other hitchhiking SNPs or MGEs that can have further consequences on bacterial fitness.

Secondly, we showed that SXT ICEs can restrict beneficial foreign DNA. Taken with previous work showing that an SXT ICE found in *V. cholerae* can restrict DNA uptake and recombination via natural transformation (70), this demonstrates that SXT ICEs can impose a barrier to genetic exchange between even closely related bacterial taxa. These types of barriers can lead to sympatric speciation, which can be revealed by further monitoring of *V. cholerae* isolates from endemic regions to track lineage shifts and correlate SXT ICE prevalence and abundance with correlated changes in other genetic diversity.

One such beneficial MGE which we found could be restricted by SXT ICEs are the phage-inducible chromosomal island-like elements, PLEs, which are another persistent feature of epidemic *V. cholerae*'s mobilome. Unlike the defense systems encoded by SXT ICEs, which collectively inhibit all circulating cholera phages, PLEs specifically inhibit the dominant lytic phage ICP1. PLEs are parasites of ICP1 and must inhibit ICP1 from packaging its genome into virions to complete their lifecycles. We uncovered here a novel mechanism that PLE deploys to target ICP1's genome and prevent packaging. PLE-encoded NixI is a nicking endonuclease that introduces numerous nicks into ICP1's genome. Such profuse nicking during ICP1's transition to rolling-circle replication (RCR) could lead to the collapse of concatemer formation, the DNA substrate required for packaging. We further uncovered a small 77-amino acid protein called StiX that abrogates NixI toxicity in a ClpXP-dependent manner, perhaps by acting as a

proteolytic adaptor leading to NixI degradation. We propose that rapid turnover of NixI may contribute to PLE's protection from NixI self-targeting. These mechanisms shed light on how PLE ensures that only its own genome can be packaged into modified ICP1 virions.

Despite these effective defenses, cholera phages persist through continued adaptation to overcome *V. cholerae*'s defense systems. In Chapter 2, we revealed how ICP1 can overcome the defenses conferred by two of the circulating SXT ICEs; *VchInd5* and *VchInd6*. ICP1 can overcome *VchInd6* through stochastic acquisition of genome modification during infection; a mechanism of escaping restriction-modification and BREX defense systems that has been described for other systems. What was more surprising was our failure to generate epigenetic escapes on *VchInd5* or *VchBan9*: two SXT ICEs that inhibit ICP1 with divergent BREX systems. It is possible that ICP1's large genome size prohibits the acquisition of sufficient methylation to subvert BREX restriction. However, it is additionally possible that these BREX systems do not function through the canonical bipartite functions of methylation and restriction, rendering escape through methylation impossible. Further work is additionally needed to understand how certain isolates ICP2 and ICP3 overcome SXT ICE-mediated defenses. Together, these novel inhibitory proteins may yield further insight into BREX function, and fitness tradeoffs encountered by phages when they encounter multiple defense systems at once.

To overcome PLE, ICP1 encodes either a CRISPR-Cas system with spacers that target PLE (34) or *Odn*, a nuclease that specifically targets PLE's origin of replication (102). These mechanisms render ICP1 insensitive to PLE, although there are likely several mechanisms employed by ICP1 to overcome PLE defenses that allow for only rare or partial escape. The work here on NixI inhibition was performed with ICP1 2006_E which was engineered to lack CRISPR-Cas machinery. Ectopic expression of NixI completely inhibits ICP1 2006_E progeny production, though other isolates of ICP1 may prove to be less sensitive to NixI nicking. Preliminary work suggests that ICP1 2011_A is not sensitive to NixI when ectopically expressed, although it is not due changes in the sequence motif recognized by NixI which persist at high numbers in this isolate's genome. It is tempting to speculate that ICP1 2011_A encodes its own *ClpXP* adaptor that targets NixI for degradation. If so, this adaptor may target other PLE-encoded proteins for degradation rendering ICP1 less sensitive to PLE overall. Further work is needed to uncover how variation in ICP1 isolates may be driven by adaptation to PLE-encoded mechanisms of inhibition. ICP1-encoded counter-adaptations may explain the variability seen between distinct PLEs and may select for replacement of one PLE with another. Intriguingly, although all PLEs encode a homolog of NixI, preliminary data suggests that PLEs 3, 4 and 5 encode NixI variants that are much less toxic when overexpressed. We propose that these less toxic homologs of NixI may be targeted for degradation by *ClpXP*, or other proteases, through inherent recognition of degron sequence, which has been lost in the NixI found in PLEs 1, 2 and 10. This can be experimentally tested by measuring the toxicity of PLEs 3, 4 and 5's NixI's in a $\Delta clpXP$ (or other protease knockout backgrounds) *V. cholerae* host and assessing whether toxicity is increased. Through tight co-evolution between NixI and *StiX*, PLE1, 2 and 10 homologs of NixI may be more stable in the absence of *StiX*, and therefore more toxic when ectopically expressed and potentially more inhibitory to ICP1. Additionally, we showed in Chapter 3 that NixI homologs are a feature of an expanded family of putative phage parasites found in non-toxigenic *V. cholerae* and other *Vibrios*, which we propose share an evolutionary origin with PLEs found in toxigenic *V. cholerae*. We noted that these PLEs lack a homolog of *StiX*, suggesting that their NixI homologs may use other mechanisms to regulate NixI protein

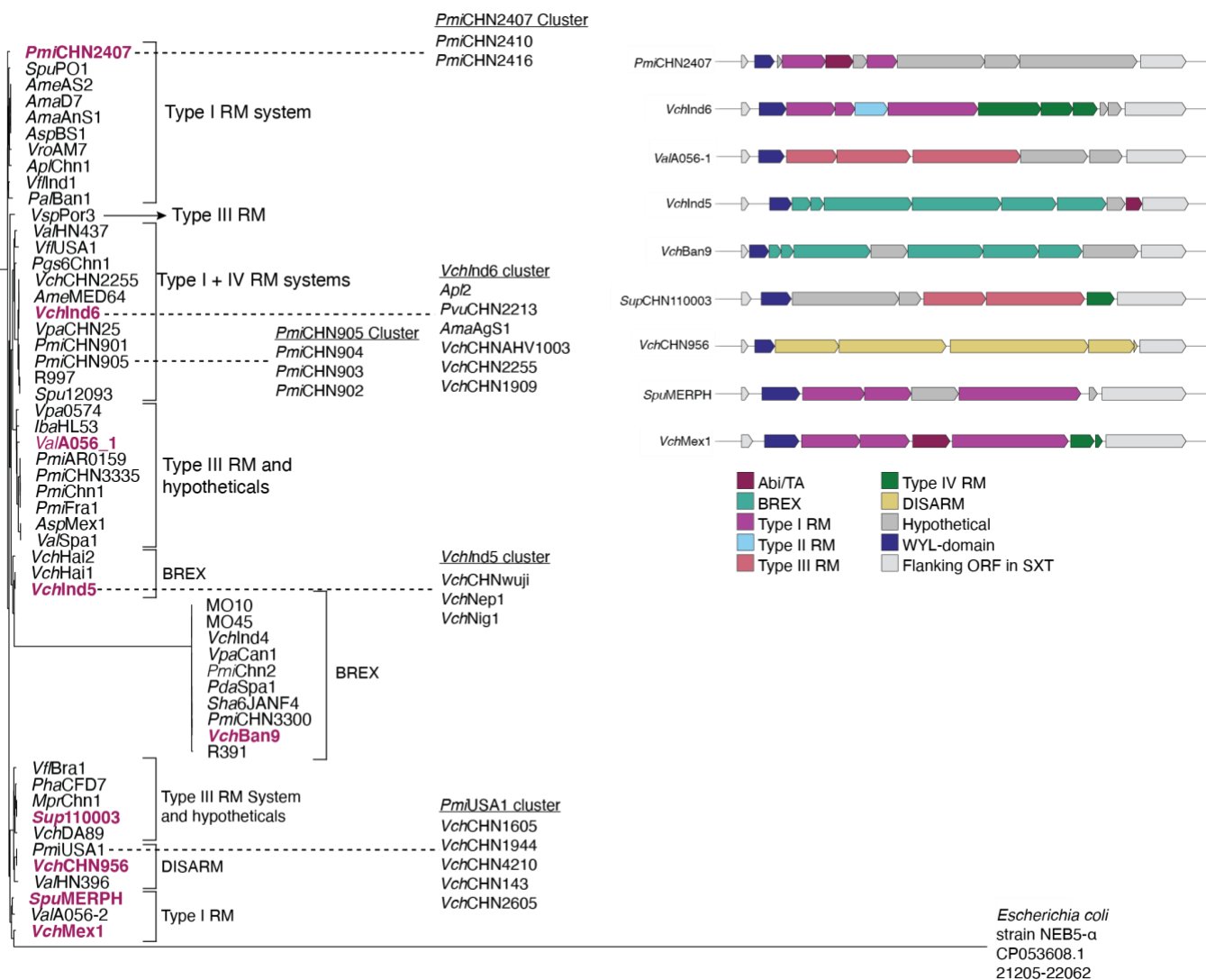
levels. Phage parasites such as PLE constitute an excellent model for addressing the role of protein turnover throughout temporally controlled phage infection cycles.

V. cholerae and its lytic phages represent an unparalleled model for uncovering the mechanisms underpinning the phage-host arms race. Recent advances in the sequencing and discovery of phage defense systems in bacterial genomes reveals the centrality of phages to bacterial evolution, as up to 10% of a bacterium's genome can be comprised of defense systems (9), though this number likely misses more nuanced mechanisms of phage evasion, such as changes to host metabolism or surface receptor masking (7). Although there are likely many more novel defense systems to be discovered and mechanistically understood, some of the most exciting discoveries in phage biology is the degree of evolutionary conservation of anti-viral strategies between prokaryotic and eukaryotic defense systems. For example, recent work has shown that bacteria may have been the ancient ancestor for eukaryotic STING-cGAS anti-viral defense (150, 151) as well as pore-forming gasdermins (152). Through mechanistic study, we may discover many more innate immune responses in eukaryotes have origins in their ancient prokaryotic ancestors (153). Further, the overlap between eukaryotic immune functions that destroy viral-infected cells with mechanisms that are used to destroy cancerous cells implies that anti-cancer mechanisms may have an origin in prokaryotic defense systems.

Additionally, the focus on phages has uncovered functions for some mysterious prokaryotic genetic modules. Retrons are enigmatic three ORF modules composed of an effector protein, reverse transcriptase and a non-coding RNA that is reverse transcribed to generate a chimeric RNA/DNA molecule (19). Together, these components cause premature cell death when a bacterium is infected, halting phage progeny production and protecting the population by limiting circulating phage titers (19, 154). Toxin-antitoxin systems likely have diverse functions in their hosts, but are now appreciated as phage defense systems (146). Upon phage infection, host transcriptional shutoff (a host-takeover strategy used by some but not all phages) results in TA system activation, leading to cell stasis or death (155, 156). These mechanisms do not appear to directly interfere with the phage lifecycle, but rather renders the host an inhospitable environment for genome replication, protein production and packaging. It is likely that in native contexts, these types of "abortive infection" systems work in concert with other defense systems, either as a last line of defense or as stopgap until a slower acting defense system, such as a naïve CRISPR-Cas system, can mount a defense and clear the infection within a cell. The field of phage biology will benefit from studying the interactions between defense systems in native contexts, to understand how cooperation between defense systems allows bacteria to keep pace with their rapidly evolving antagonists.

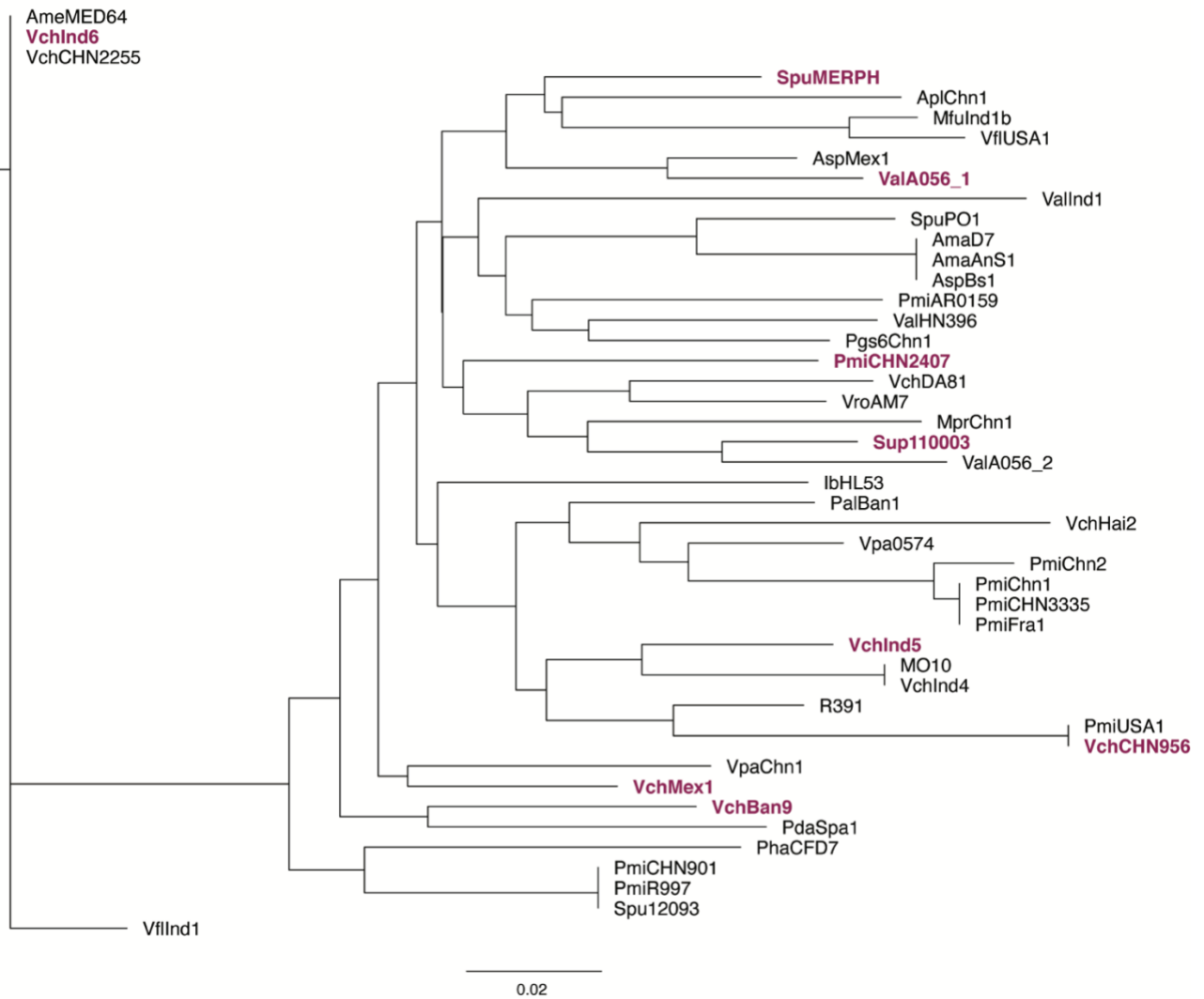
Supplementary Materials

A

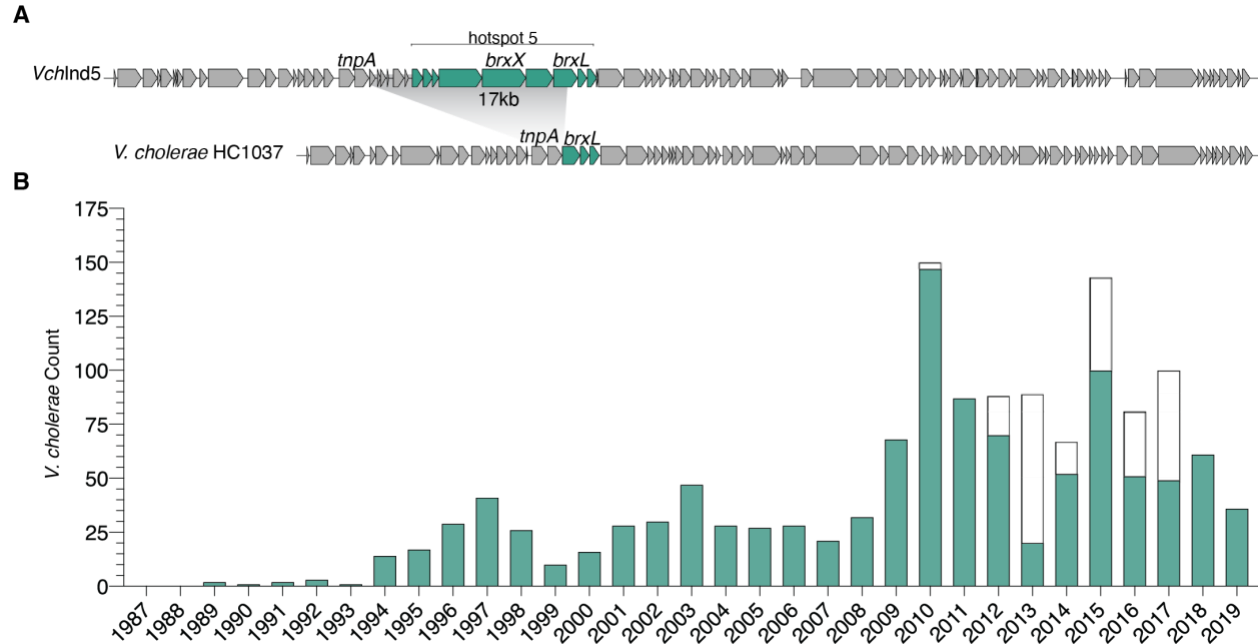


3.0

B

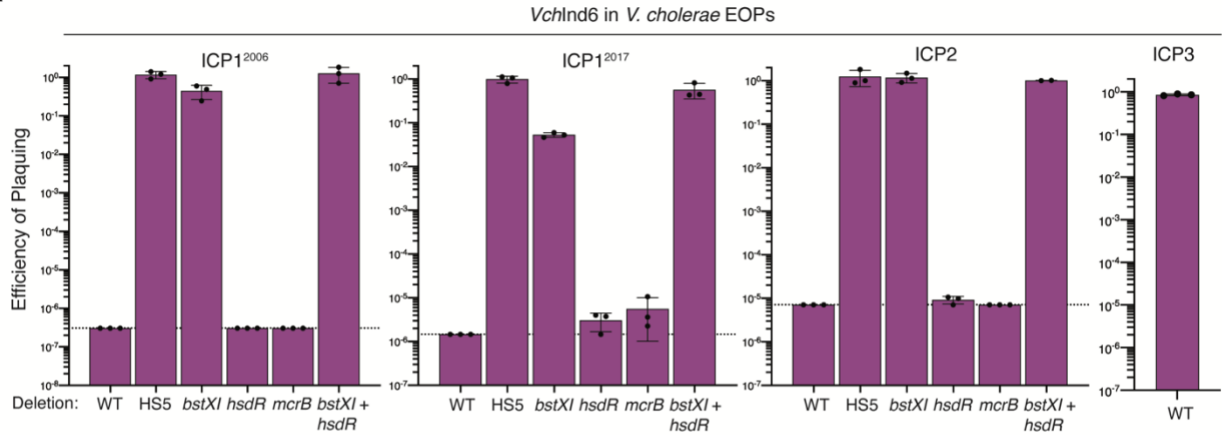


Supplementary Figure 1.1. Hotspot 5 in SXT ICEs encode a conserved WYL-domain protein. (A) Neighbor-joining phylogenetic tree from a MUSCLE alignment of the nucleotide sequence for the WYL-domain containing gene encoded in each SXT ICE, using a WYL-domain protein from *Escherichia coli* as an outgroup. Clusters of SXT ICEs with identical WYL-domain proteins are shown with a dashed line to the right, modified from Bioteau et al. 2018 (29). The predicted anti-phage system encoded by hotspot 5 is indicated beside the bracket for each group. Gene maps (not to scale) for representative SXT ICEs (bolded in red) from each node are depicted next to the tree and shows the upstream hypothetical s025 and downstream conserved *traI* genes as flanking genes in hotspot 5. (B) Neighbor-joining phylogenetic tree from a MUSCLE alignment of the nucleotide sequence for conserved downstream gene *traI* showing no relationship between hotspot 5 anti-phage type system and similarity.

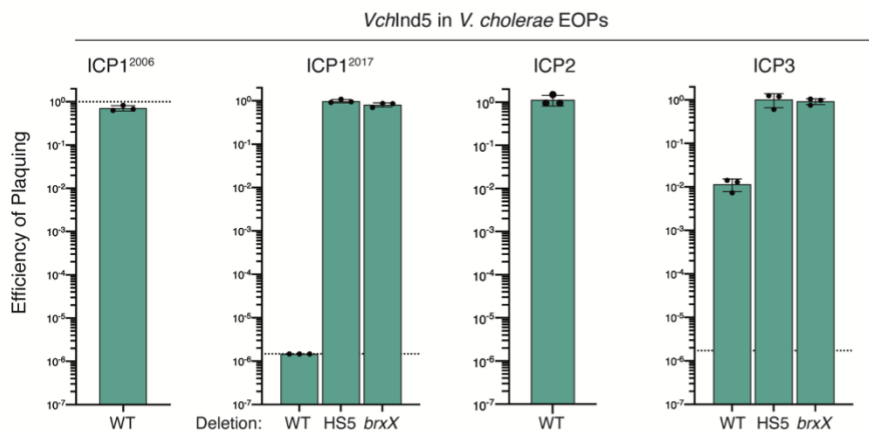


Supplementary Figure 1.2. Prevalence of ICEVchInd5 hotspot 5 deletions in toxigenic O1 *V. cholerae*. (A) Schematic comparing wildtype ICEVchInd5 (top) indicating hotspot 5 in teal compared to ICEVchInd5 in *V. cholerae* isolate HC1037 (accession CP026647) containing the 17kb deletion spanning most of hotspot 5. (B) The abundance of ICEVchInd5 in 1373 *V. cholerae* isolates (total bar height) showing a total of 229 isolates identified as lacking hotspot 5 from the SXT ICE BLASTn analysis indicated as the portion of the bar shown in white. A subset of 10 of the 229 isolates were analyzed and were found to contain the deletion as in A. The remaining isolates were identified as having ICEVchInd5 SXT yet lacking hotspot 5, but the exact nature of the deletions was not confirmed for all isolates. Of the 229 isolates lacking hotspot 5, 222 were isolated in Haiti, 3 isolates were from Nepal (in 2010) and 4 isolates were from Mexico (from 2013).

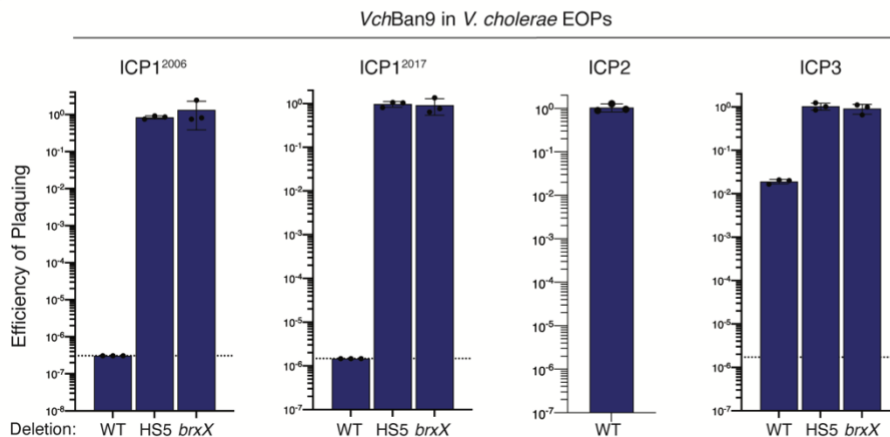
A



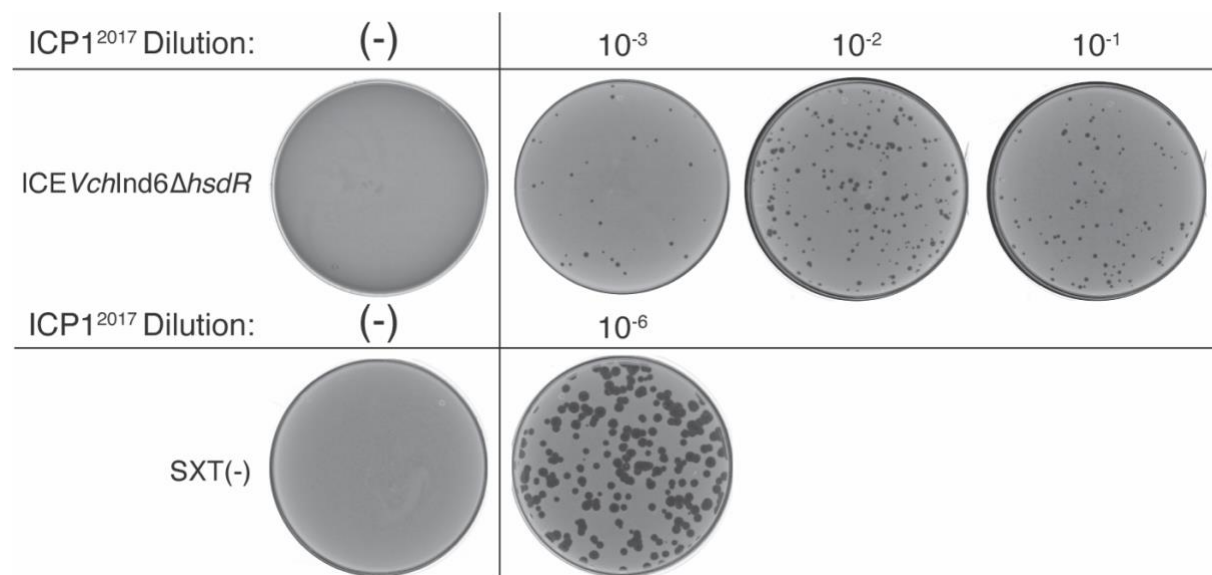
B



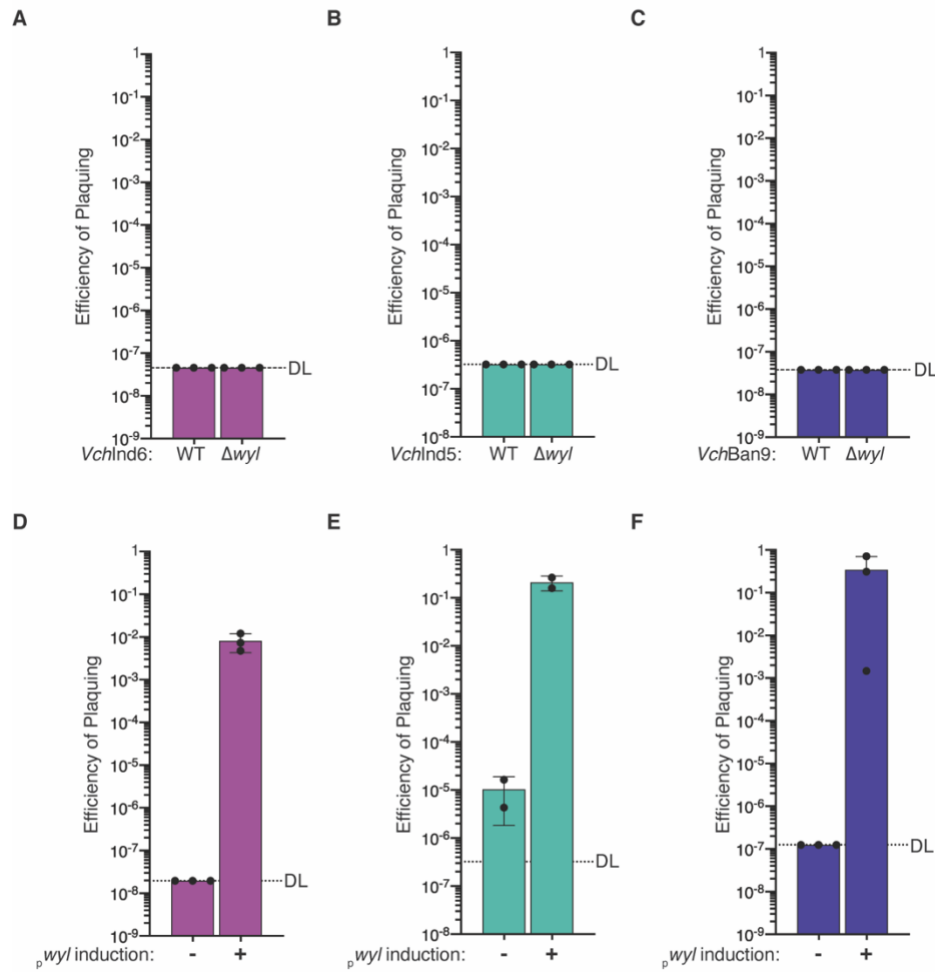
C



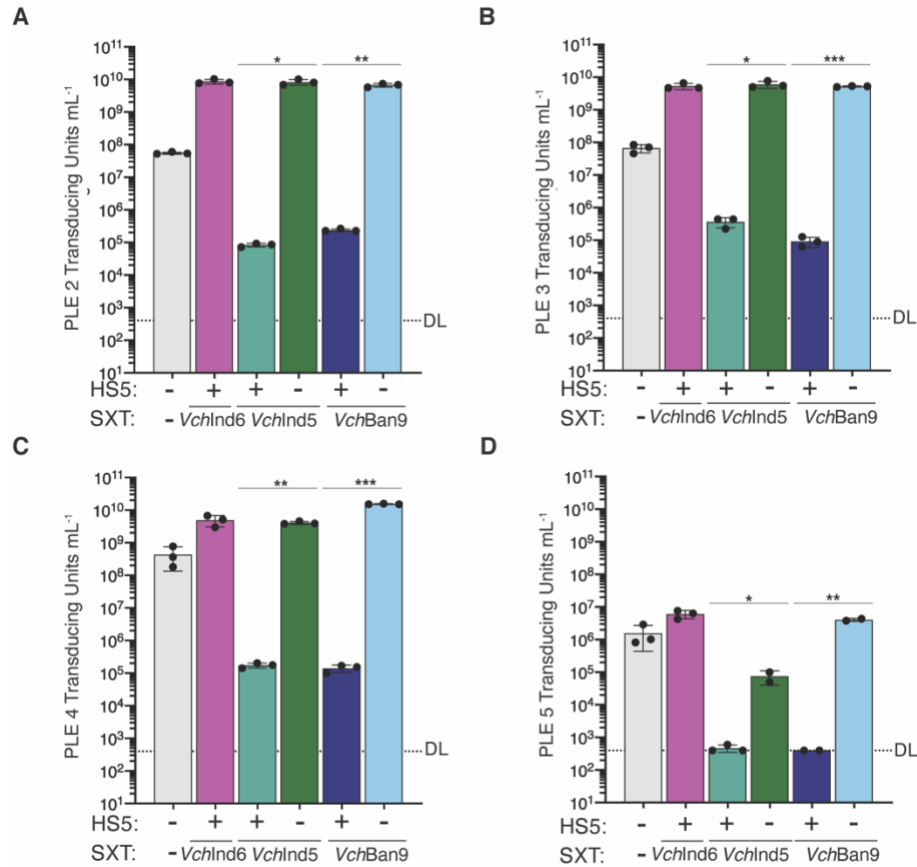
Supplementary Figure 1.3. SXT ICEs confer protection against lytic vibriophages. Source data for Figure 2B in which wild type and SXT ICE knockouts were conjugated into *V. cholerae* E7946 and the efficiency of plaquing (EOP) was calculated for the lytic phages ICP1²⁰⁰⁶, ICP1²⁰¹⁷, ICP2²⁰⁰⁴, and ICP3²⁰⁰⁷. (A) EOPs for phages on ICEVchInd6 *V. cholerae*. (B) EOPs for phages on ICEVchInd5 *V. cholerae*. (C) EOPs for phages on ICEVchBan9 *V. cholerae*. WT indicates the wild-type SXT ICE, HS5 is used to indicate the hotspot 5 deletion. Dots represent individual biological replicates; bar height is the mean; error bars display the standard deviation and the dashed line indicates either the limit of detection.



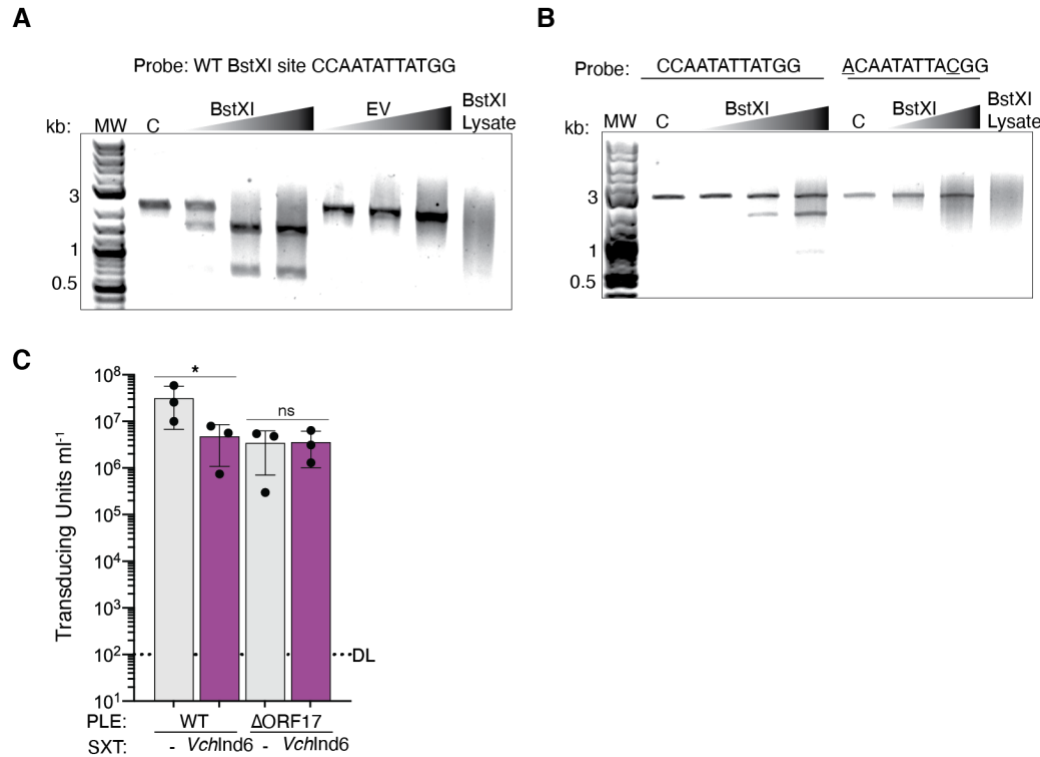
Supplementary Figure 1.4. ICEVchInd6 *hsdR* is necessary for full inhibition of ICP1²⁰¹⁷. Plaque assays of ICP1²⁰¹⁷ on *V. cholerae* ICEVchInd6Δ*hsdR* (top) and SXT(-) (bottom), bacterial lawns in gray and zones of killing shown in black. Uninfected, far left. Dilution series of ICP1²⁰¹⁷ from 10⁻³ up to 10⁻¹ (top) and 10⁻⁶ on SXT(-). Results are representative of assays performed in triplicate.



Supplementary Figure 1.5. The conserved WYL-domain protein may act as a repressor for hotspot 5-mediated resistance to ICP1. Top panels: Efficiency of plaquing of ICP1 on wild type (WT) and each SXT derivative with the WYL-domain protein deleted, from (A) ICEVchInd6 (purple), (B) ICEVchInd5 (teal) and (C) ICEVchBan9 (blue). ICP1²⁰⁰⁶ was assayed on ICEVchInd6 and ICEVchBan9 and ICP1²⁰¹⁷ was assayed on ICEVchInd5. (D-F): Efficiency of plaquing of ICP1 (isolate tested same as in top panel) in an SXT ICE(+) host with ectopic expression of the WYL-domain protein. The WYL-domain protein restores infection by ICP1 (isolate tested same as in top panel) when induced prior to performing plaque assays (indicated by the “+” below the x-axis) compared to the same background uninduced (indicated by the “-” sign). Dots represent independent biological replicates; bar height is the mean; error bars display the standard deviation. DL indicates the limit of detection.

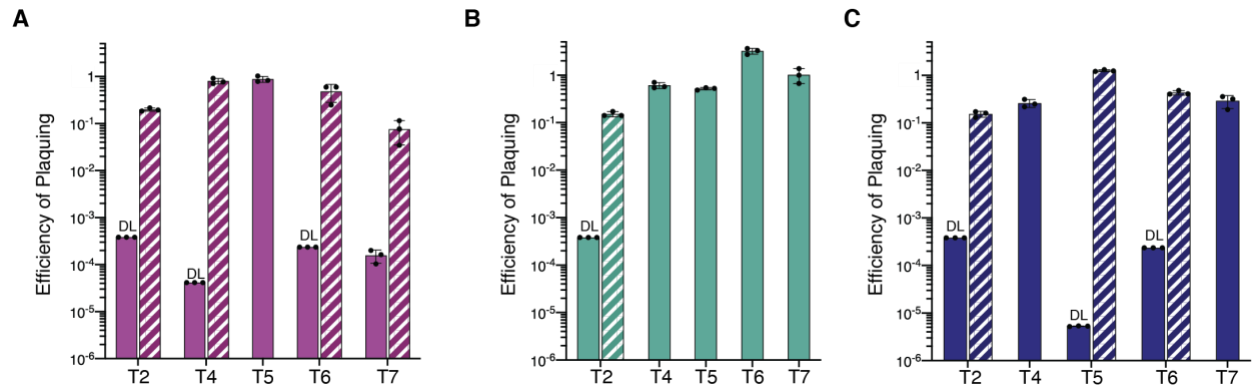


Supplementary Figure 1.6. SXT ICEs limit acquisition of PLEs, mobile phage satellites that provide anti-phage activity. Transductions of (A) PLE 2 (B) PLE 3 (C) PLE 4 and (D) PLE 5 into *V. cholerae* E7946 SXT(-) and E7946 harboring SXT ICEs and the derivative Δ hotspot 5 (HS5) recipients. ICEVchInd6 did not inhibit PLEs 2-5 so the ICEVchInd6 Δ HS5 strain was not tested. Where transduction into the Δ hotspot 5 recipient was greater than the SXT(-) recipient, we anticipate this is due to the spontaneous appearance of O-antigen mutants in the culture, which has been shown previously to be required in the recipient for PLE transduction (20). Dashed lines indicate the limit of detection (DL), dots represent independent biological replicates, bar height is the mean; error bars display the standard deviation. Asterisks denote significance of relevant comparisons, where no asterisks are shown, statistics were not performed (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$, paired t test)

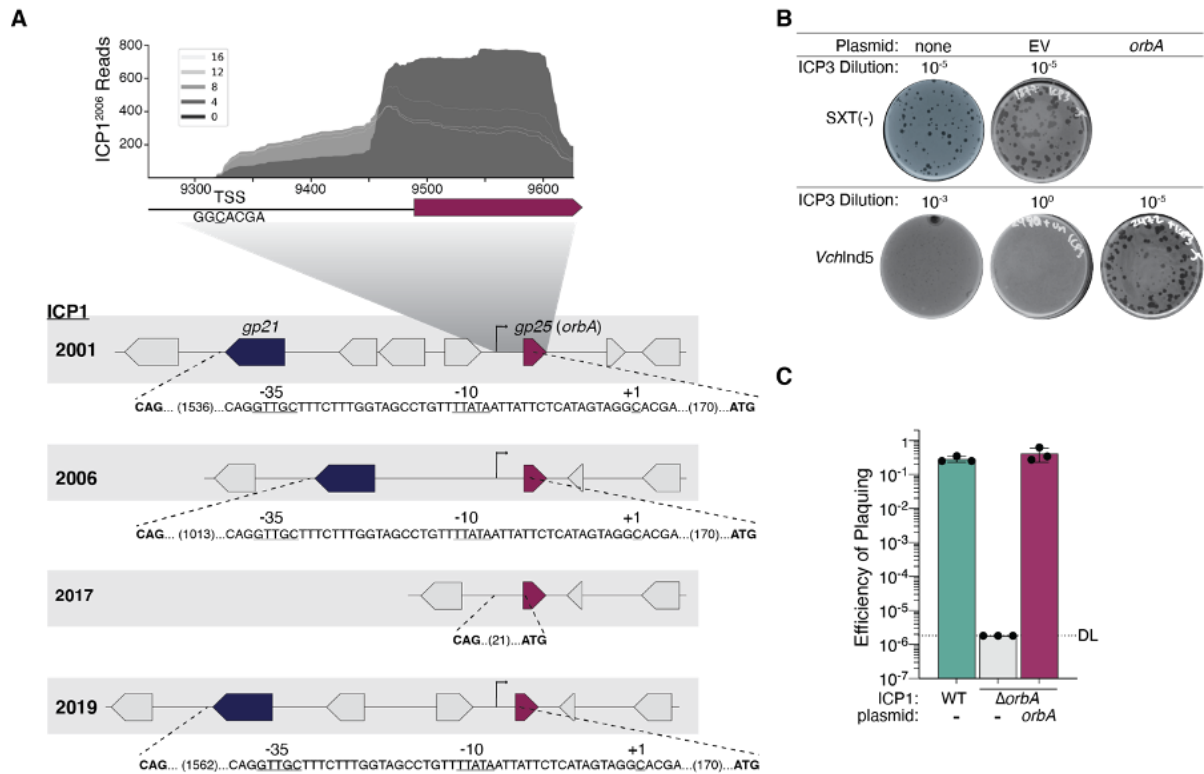


Supplementary Figure 1.7. BstXI from ICEVchInd6 cleaves the predicted recognition motif

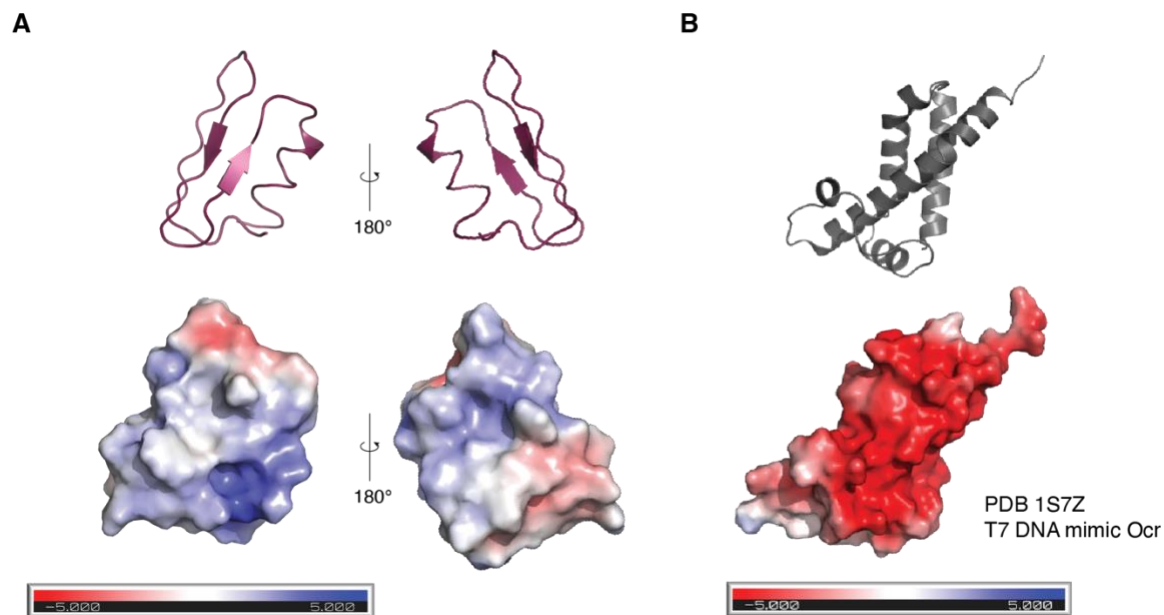
CCANNNNNTGG. (A) Lysate containing BstXI, but not empty vector lysate (EV), cleave a DNA probe with the predicted recognition motif generating cleavage products of predicted size. Shaded triangles above the agarose gel indicate increasing volume of lysates (2µl, 5µl and 10µl) in 20µl total reaction volumes containing a probe amplified from PLE 1's ORF17 region. (B) Single point mutations in the recognition motif abolishes cutting by BstXI. Shaded triangles above the gel show increasing volume (2µl, 5µl and 10µl) of BstXI lysate with the native probe in lanes 3-5, and in lanes 7-8 show 5µl and 10µl of BstXI lysate with the mutated probe. For A and B, the final lanes are 10µl of BstXI lysate with no added DNA probe; molecular weight ladder (MW) is in lane one, and the probe alone is indicated by "C". (C) Wild-type (WT) PLE 1 (left bars) and PLE 1ΔORF17 (righthand bars) transduction into an SXT(-) (in grey) and ICEVchInd6(+) (purple) recipient showing that deletion of ORF17, containing the recognition site for ICEVchInd6(+)’s BstXI, allows for equivalent transduction into both recipients. Dashed line indicates the limit of detection (DL), dots represent independent biological replicates, bar height is the mean; error bars display the standard deviation. Asterisks denote significance (ns>0.05; *P <0.05, one-way ANOVA, Fisher’s LSD).



Supplementary Figure 1.8. SXT ICEs confer protection against lytic phage in *E. coli*. Source data for Figure 3B. Wild type SXT ICEs (solid bars) and Δ hotspot 5 knockouts (patterned bars) were conjugated into *E. coli* MG1655 and the efficiency of plaquing was calculated for the lytic phages T2, T4, T5, T6 and T7. (A) ICEVchInd6 and ICEVchInd6 Δ H5S5. (B) ICEVchInd5 and ICEVchInd5 Δ H5S5. (C) ICEVchBan9 and ICEVchBan9 Δ H5S5. Where plaques are at the limit of detection, this is indicated by a “DL” over that bar. Dots represent independent biological replicates; bar height is the mean; error bars display the standard deviation.

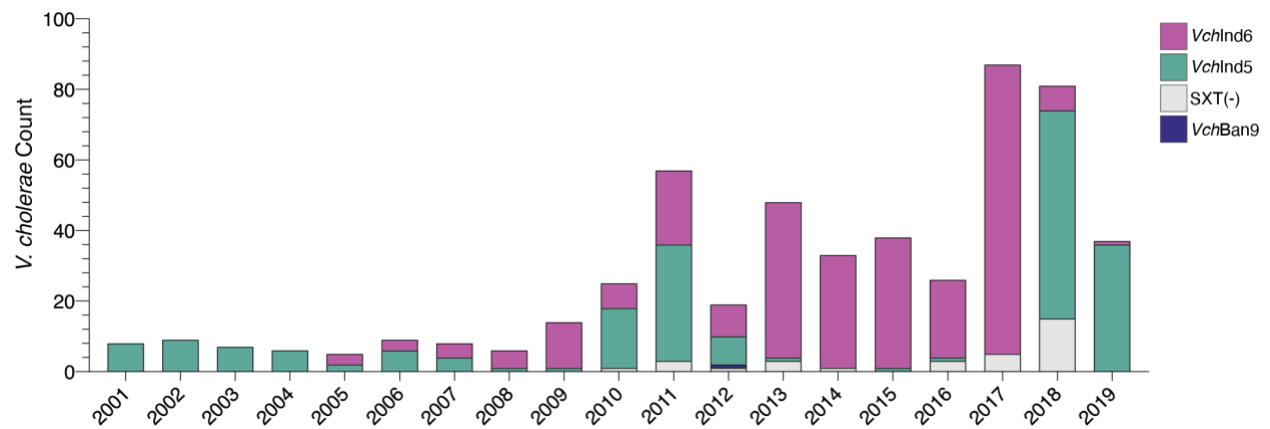


Supplementary Figure 1.9 Expression of ICP1-encoded *gp25* (*orbA*) is necessary to overcome ICEV*chInd5* immunity. (A) Top: RNA-seq of ICP1²⁰⁰⁶ mapping reads onto the region encompassing *gp25* and its upstream region were used to identify the likely transcriptional start site (TSS). Gene maps of ICP1 isolates used in Figure 4C are shown above the sequence upstream of *gp25* in the dashed-line inset for each phage, showing the -35 and -10 promoter elements and +1 TSS. The bracketed numbers indicate the number of nucleotides between sequences. Nucleotides in bold on left show the beginning of conserved flanking sequence, and the bold ATG on the right shows the start codon for *gp25*. (B) Representative images of ICP3 plaque phenotype on ICEV*chInd5* showing smaller plaques compared to SXT(-). Maintaining an empty vector plasmid (EV) rendered plaques not visible, though ectopic expression of *orbA* restored plaque size. Bacterial lawns are in gray and zones of killing shown in black. (C) A clean deletion of *orbA* in ICP1²⁰⁰⁶ abolishes plaquing on ICEV*chInd5*, and complementation of *orbA* from a plasmid restores plaquing. Dots represent individual biological replicates, bar height is the average; error bars display the standard deviation, dashed line shows the limit of detection (DL).



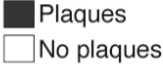
Supplementary Figure 1.10. ICP1 encoded OrbA predicted structure and surface electrostatic potential map.

(A) Predicted structure of OrbA from i-TASSER (71) showing 180 degree rotation of around the vertical axis showing predicted beta sheets and helix, with APBS electrostatics overlayed indicating negative charged surfaces with red and positively charged surfaces in blue. (B) Solved crystal structure of the known anti-BREX protein T7 DNA mimic Ocr (PDB 1S7Z) with electrostatics overlayed on the structure.



Supplementary Figure 1.11. Prevalence of SXT ICEs in toxigenic O1 *V. cholerae* from Bangladesh. Sampling and whole genome sequencing of 523 *V. cholerae* isolates from Bangladesh, depicting isolate counts (y-axis) by year (x-axis), highlighting sporadic occurrences of SXT(-), ICE*VchBan9*(+) and fluctuating dominance of SXT ICEs, from ICE*VchInd5*(+) to ICE*VchInd6*(+) followed by a subsequent return to ICE*VchInd5*(+) *V. cholerae*.

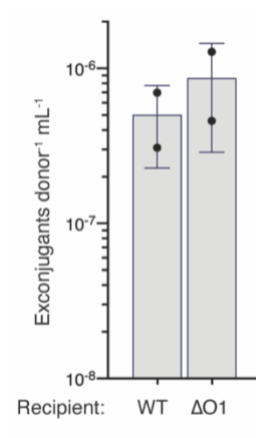
			ICP1		Podoviridae	
			2006	2017	ICP2	ICP3
SXT(-) Patient Isolate	Patient	BFS	Date			
	DCP-53	BFS618	2017.06.04			
	DCP-66	BFS626	2017.12.20			
	DCP-68	BFS292*	2018.02.12			
	DCP-68	BFS293*	2018.02.12			
	DCP-68	BFS474*	2018.02.12			
	DCP-73	BFS11	2018.02.18			
	DCP-73	BFS12	2018.02.18			
	DCP-74	BFS476*	2018.02.20			
	DCP-81	BFS478*	2018.04.09			
	DCP-90	BFS482*	2018.05.07			



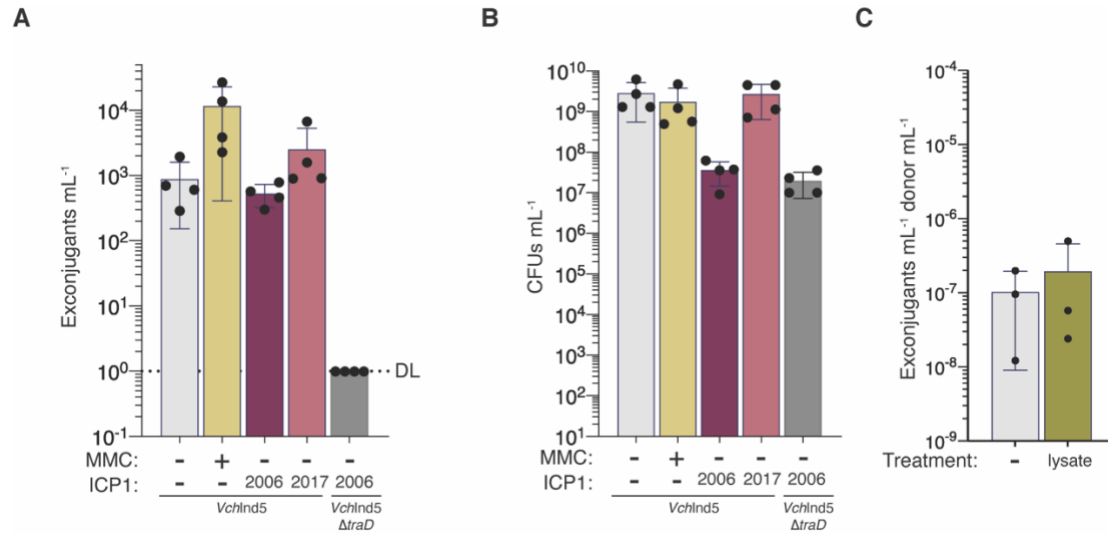
Plaques

No plaques

Supplementary Figure 1.12. SXT ICE(-) *V. cholerae* isolates do not possess additional anti-phage defense systems. Matrix summarizing the results of spot assays of ICP1 (two isolates, from 2006 and 2017), ICP2 and ICP3 on *V. cholerae* patient isolates (where BFS indicates "Bacteria From Stool" number) identified as being SXT(-) from whole genome sequencing. Isolates with an asterisk were resistant to ICP2 and possess a mutation in ICP2's receptor (OmpU) previously shown to block ICP2 infection (9). Spot assays were repeated two independent times, and plaquing is defined as a full restoration in infection.



Supplementary Figure 1.13. SXT ICE conjugation does not require the O1-antigen. Exconjugants ml⁻¹ normalized to the number of ICEVchInd5 donor cells (KL571) into *V. cholerae* E7946 recipient (KS2418) or a *V. cholerae* donor lacking the ICP1 receptor, O1 antigen (KL322=E7946 ΔVC0249). Dots represent individual biological replicates; bar height is the average; error bars display the standard deviation, means were not significantly different (ns<0.05, paired t test).



Supplementary Figure 1.14. ICP1 stimulated SXT ICE conjugation results in cell death but equivalent levels of conjugation and is not driven by increased nutrient availability. (A) Raw numbers of exconjugants obtained following conjugation with donor ICEVchInd5(+) *V. cholerae* treated with mitomycin C (MMC), or infected by ICP1²⁰⁰⁶ or ICP1²⁰¹⁷ prior to conjugation compared to an untreated control. The dashed line indicates the limit of detection (DL). (B) Enumeration of surviving ICEVchInd5(+) *V. cholerae* donors treated with mitomycin C (MMC), or infected by ICP1²⁰⁰⁶ or ICP1²⁰¹⁷ following conjugation compared to an untreated control. (C) Addition of lysate from mid-log *V. cholerae* that was chloroform treated then lysed by sonication and passed through a 100kDa molecular weight cut-off filter was used to resuspend the conjugation pellet as well as to soak the filter on which the conjugation takes place, and was not found to increase conjugation (ns<0.05, t test). For all plots, dots represent individual biological replicates, bar height is the average; error bars display the standard deviation.

Supplementary Table 1.1. Oligos used in this study

Name	Sequence (5'-3')	Purpose
KS1114	ACGTACAGCGTAAGACATTGG	ICP2 screening
KS1119	GAACAAATCTGTGCTCGTATCG	ICP2 screening
SGH96	ATGGCAGCTTTTAAGAAACGTACACAAACC	ICP3 screening
SGH97	TTACTTCACATTAGCCAATAGACCATG	ICP3 screening
KS923	GAGATAAAATACCGTCCTTGGG	ICP1 screening
KS924	TGCATCTTTAACCCACAACC	ICP1 screening
ACM279	TCTTGCCGATGCAGTGAAAC	ICP1 screening
ZB61	TCTGTGCGGTGTGGGTAATA	ICP1 screening
TU18gp5L	ATTGTCGAGTGGGACAAAGG	ICP3 screening
TU18gp5R	ACCAACTCGACGCATAGCTT	ICP3 screening

Supplementary Table 1.2. Presence of SXT ICEVchInd6 hotspot 5 contents outside of *V. cholerae*.

Taxa	Strain	Accession	Query Coverage	Percent Identity
<i>Proteus vulgaris</i>	08MAS2213	KX243403.1	100%	99.99
<i>Alteromonas mediterranea</i>	MED64	CP004848.1	100%	99.99
<i>Actinobacillus pleuropneumoniae</i> (ICEApl2)	MIDG3553	MF187965.1	100%	99.79
<i>Shewanella putrefaciens</i>	NCTC12093	LR134303.1	99%	99.50%
<i>Proteus mirabilis</i> pR997	R997	KY433363.1	99%	99.49
<i>Providencia rettgeri</i>	2055	CP059345.1	99	99.49
<i>Providencia rettgeri</i>	RF14-2 ICEPreChnRF14-2	MT219827.1	100	99.90
<i>Proteus mirabilis</i>	MD20140905 MD20140904 MD20140903 MD20140902 S012	KX243412.1 KX243411.1 KX243410.1 KX243409.1 CP062146.1	99	99.49
<i>Proteus mirabilis</i>	MD20140901	KX243408.1	99	99.73
<i>Pseudoalteromonas</i> sp. (ICEPspSpa1)	4mar partial ICEPspSpa1	HG794401.1	79	96.45

Supplementary Table 1.3. Presence of SXT ICE*Vch*Ind5 hotspot 5 contents outside of *V. cholerae*.

Taxa	Strain	Accession	Query Coverage	Percent Identity
<i>Proteus mira</i>	AR_0155	CP021694.1	93	96.24
<i>Shewanella</i> s	FDAARGOS_354	CP022089.2	82	95.59

Supplementary Table 1.4. Presence of SXT ICEVchBan9 hotspot 5 contents outside of *V. cholerae*.

Taxa	Strain	Accession	Query Coverage	Percent Identity
<i>Proteus mirabilis</i>	MPE5139	CP053684.1	100	100
<i>Proteus mirabilis</i>	Pm15C1; ICEPmiChn2	KX268685.1	100	100
<i>Proteus mirabilis</i>	TJ3300	KX243415.1	100	100
<i>Vibrio parahaemolyticus</i>	PB1937	CP022243.1	91	98.28
<i>Escherichia coli</i>	ISO4 plasmid pLS04_68	HG963476.1	78	97.86
<i>Proteus mirabilis</i>	MPE0027	CP053683.1	93	97.40
<i>Escherichia coli</i>	Cq9	CP031546.1	93	97.42
<i>Proteus</i> sp. <i>Proteus cibarius</i>	ZN5	CP047639.1	78	97.83
<i>Proteus mirabilis</i>	ICEPmiChn2	KY437726.1	78	97.83
<i>Klebsiella michiganensis</i>	RHBSTW-00676	CP055985.1	79	96.97
<i>Escherichia coli</i>	PCN061 plasmid PCN061p6	CP006642.1	78	97.55
<i>Escherichia coli</i>	AR_0011	CP024857.1	79	97.18
<i>Vibrio anguillarum</i>	CNEVA NB11008	CP022103.1	79	96.43
<i>Salmonella enterica</i>	R19_2939 Plasmid pR19.2939_83K	CP046430.1	79	96.45
<i>Escherichia coli</i>	2248, plasmid pCTXM-224	MG836696.1	79	96.45
<i>Escherichia fergusonii</i>	ATCC 35469, plasmid pEFER	CU928144.1	78	97.15
<i>Photobacterium damsela</i>	Subsp. Piscicida ICEPdaSpa1 isolate PC554.2	AJ870986.2	81	97.86
<i>Vibrio parahaemolyticus</i>	S107	CP028481.1	81	97.25
<i>Vibrio diabolicus</i>	FDAARGOS_96	CP014094.1	68	97.94
<i>Acinetobacter schindleri</i>	SGAir0122	CP025618.2	70	93.34

Supplementary Table 2.1. Strains used in this study

Strain	Description*	Source
KDS6	<i>V. cholerae</i> E7946	Lab collection
KDS103	<i>V. cholerae</i> E7946 containing PLE1 Kan ^R	(32)
KDS36	<i>V. cholerae</i> E7946 containing PLE1	(32)
KDS162	<i>V. cholerae</i> E7946 PLE1 $\Delta nixI::frr$, Kan ^R	(100)
KL228-229	<i>V. cholerae</i> E7946 PLE1 $\Delta nixI$ pKL06: P _{tac} Empty Vector, Kan ^R , Cm ^R	This study
KL230-231	<i>V. cholerae</i> E7946 PLE1 $\Delta nixI$ pKL06: P _{tac} nixI, Kan ^R , Cm ^R	This study
KDS196	<i>V. cholerae</i> E7946 pKL06 empty vector Cm ^R	(100)
KDS211	<i>V. cholerae</i> E7946 pKL06: P _{tac} nixI Cm ^R	This study
KL160-161	<i>V. cholerae</i> E7946 pKL06: P _{tac} nixI ^{N95A} Cm ^R	This study
KL735-736	<i>V. cholerae</i> E7946 PLE1 with the region from <i>gp121</i> cloned 3' of last ORF, Kan ^R	This study
KL591-592	<i>V. cholerae</i> E7946 pKL06: P _{tac} nixI homolog from <i>V. cholerae</i> YB2_A06, Cm ^R	This study
KL593-594	<i>V. cholerae</i> E7946 pKL06: P _{tac} nixI homolog from <i>V. parahaemolyticus</i> s042, Cm ^R	Thus study
KL249-250	<i>V. cholerae</i> E7946 midiPLE, $\Delta lacZ::EV$, pKL06 P _{tac} -repA, Kan ^R , Spec ^R , Cm ^R (RepA plasmid expression construct in strain with midiPLE and empty vector construct in lacZ locus)	This study
KL251-252	<i>V. cholerae</i> E7946 midiPLE, $\Delta lacZ::P_{tac}-nixI$, pKL06 P _{tac} -repA, Kan ^R , Spec ^R , Cm ^R (RepA plasmid expression construct in strain with midiPLE and inducible <i>nixI</i> in lacZ locus)	This study
KL133	<i>E. coli</i> BL21 pLysS with IPTG inducible His-SUMO- <i>nixI</i> from PLE1, Kan ^R for expression vector, Cm ^R for pLys vector (purification of NixI)	This study
KL171	<i>E. coli</i> BL21 His-SUMO- <i>nixI</i> ^{N95A} , Kan ^R (purification of catalytically inactive NixI)	This study
KL285	<i>E. coli</i> pEmpty Vector (pUC19 derivative plasmid) for empty vector control for nuclease assays, Cm ^R	This study
KL286	<i>E. coli</i> pGp121 (pUC19 derivative plasmid), <i>gp121</i> region from ICP1 2006_E, Cm ^R	This study
KL287	<i>E. coli</i> pPLE ori (pUC19 derivative plasmid), containing PLE1's putative <i>pac</i> site, Cm ^R	This study
KL645	<i>E. coli</i> pGp121 (pUC19 derivative plasmid), <i>gp156</i> region from ICP1 2006_E, Cm ^R	This study
KL804	<i>E. coli</i> pGp156* (pUC19 derivative plasmid), <i>gp156*</i> , consensus motif changed from <u>GTAATCTT</u> to <u>TGTATAGT</u> , Cm ^R	This study

KL814	<i>E. coli</i> pMotif (pUC19 derivative plasmid), motif GTAATCTT only, Cm ^R	This study
KL812	<i>E. coli</i> pPLE2 (pUC19 derivative plasmid), PLE2 region, Cm ^R	This study
ICP1	ICP1_2006_E ΔCRISPR ΔCas2-3	(82)

* Kan^R = Kanamycin resistance cassette, Spec^R = Spectinomycin resistance cassette, Cm^R=Chloramphenicol resistance cassette

Supplementary Table 2.2. Oligos used in this study

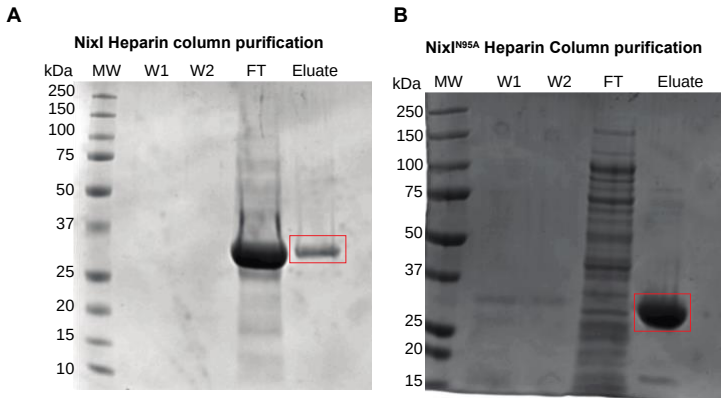
Primer	Sequence	Purpose	Source
zac14	AGGGTTTGAGTGCGATTACG	qPCR PLE	(32)
zac15	TGAGGTTTTACCACCTTTTGC	qPCR PLE	(32)
zac68	CTGAATCGCCCTACCCGTAC	qPCR ICP1	(32)
zac69	GTGAACCAACCTTTGTCGCC	qPCR ICP1	(32)
KL119	cgaacagattggaggtATGCCAACACAAAAGACGT	<i>nixI</i> with pSUMO overhangs	This study
KL120	tggtggtggtgctcgaTTTAAATTTCTAGCCGATCT	<i>nixI</i> with pSUMO overhangs	This study
KL170	CAAGAGgccTTAAAAGAACTTCTAATTTTTACTT AG	<i>nixI</i> ^{N95A} point mutation	This study
KL171	CTTTTAAggcCTCTTGGTGACTACACCAC	<i>nixI</i> ^{N95A} point mutation	This study
KL226	CTATGACCATGATTACGCCAGTCGATGTTCTAAA GGGTATCG	<i>gp121</i> overhangs to pKL285	This study
KL227	GTTGTAAAACGACGGCAGCCTCTGATAGTGCTT CGT	<i>gp121</i> overhangs to pKL285	This study
KL233	CTATGACCATGATTACGCCAAAGTTACCTCAGTG TCGTTAG	<i>gp156</i> overhangs to pKL285	This study
KL234	GTTGTAAAACGACGGCCTCATCCATCCACACAA ACCT	<i>gp156</i> overhangs to pKL285	This study
KL224	GACGTTGTAAAACGACGGCCTATATTTGAGTAT AAGTGTGGTGC	PLE's ori overhangs to pKL285	This study

KL225	GCTATGACCATGATTACGCCACTGGAAATAATG ACAATTGAGATTGC	PLE's ori overhangs to pKL285	This study
KL459	GTAATCTTGCCGTCGTTTTACAACGTCG	Motif overhangs to pKL285	This study
KL460	GTAAAACGACGGCAAGATTACTGGCGTAATCAT GGTCATAGC	Motif overhangs to pKL285	This study
KL449	ATGACCATGATTACGCCAgTGACAACACCAACT AATACAGAAGTAAAAAC	PLE2 overhangs to pKL285	This study
KL450	GACGTTGTAAAACGACGGCgTTCCTAGTTTTTGC CATCACAGC	PLE2 overhangs to pKL285	This study
KL458	GCCACCATAATACTTTTGCAATG	100bp EMSA probe of <i>gp156</i>	This study
ACM12 6	GGAGGATGATGAAATGAACG	100bp EMSA probe of <i>gp156</i>	This study
KS951	CTGCTAAGGAGGTAACAACAAGATGCCAACACA AAAGACGTTAGACC	<i>nixI</i> pKL06 overhangs	This study
KS962	ctctcatccgcaaaacagcTTAAATTTCTAGCCGATCTATT TTAAATCG	<i>nixI</i> pKL06 overhangs	This study
KL398	GGTGAACGCTCTCCTGAGTGATGGAAGGGCTAA CGGTTATTTTCC	PLE1 overhangs <i>gp121</i> region	This study
KL399	CTCATATATGGTTAATCTTGTCATCTCCGTGATT ACCTTGGG	PLE1 overhangs <i>gp121</i> region	This study

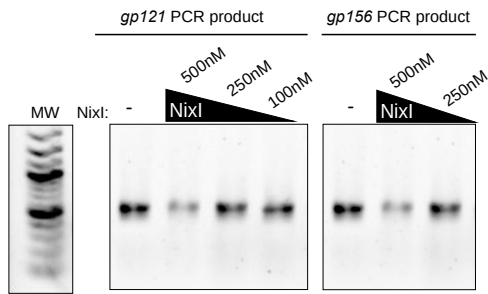
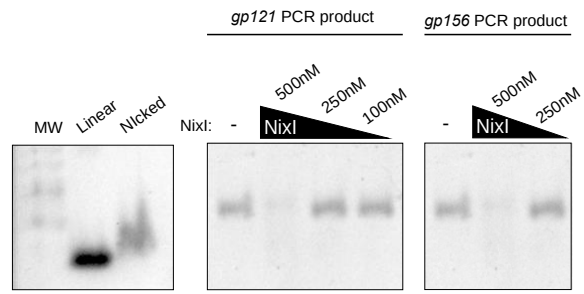
Supplementary Table 2.3. Accession numbers/sequences for NixI

NixI Strain Background	NixI Accession Number/Sequence
<i>V. cholerae</i> PLE1	WP_032468646
<i>V. cholerae</i> PLE3	MF176135.1

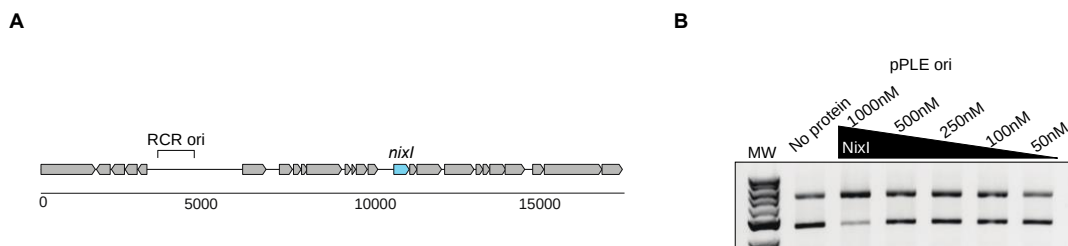
<i>V. cholerae</i> PLE4	CP001486.1
<i>V. cholerae</i> PLE5	WP_001147214.1
<i>V. cholerae</i> PLE6	MPTQKTLDRAAKINCTEAQLDKVYNQFNNIKYRCGKGGRYN NAKFLLGDFDTFIEWIASQGGYRQSPHGHAYHVSRI RDQGD YAYNNIRWCSHQENLKETS NFYLVTDLKSSRQELFQGNLAD KLKELFNVEATSGAYAASKTG NLYLGRFKIDRLEI
<i>V. cholerae</i> PLE7	MPTQKTLDRAAKINCTVKQLQGVYAQYKNIKYRCGKGGRYN NAKFLLGTFDDFIEWIESQGGFKTSPHGHAYHVSRI RDQGDY AYNNIRWCSHQENLKETANFYLVTDLKSSRQELFQGNLADK LKELFNVEATSGAYTASKTG NLYLGRFKIDKLSI
<i>V. cholerae</i> PLE9	MPTQKMLDRAAKINCTVKQLRGVYAQYKNIKYRCGKGGRY NNAKFLLGDFDTFIEWIASQGGYRQSPHGHAYHVSRI RDQG DYAYNNIRWCSHQENLKETS NFYLVTDLKSSRQELFQGNLA DKLKELFNVEATSGAYTASKTG NLYLGRFKIDRLEI
<i>V. parahaemolyticus</i> CDCK4858	WP_069547983.1
<i>V. parahaemolyticus</i> S106	WP_025626793.1
<i>V. alginolyticus</i> ANC4	WP_025626793.1
<i>V. casei</i> DSM22364-1	WP_089110726.1
<i>V. casei</i> DSM22364-2	WP_110957831.1
<i>V. parahaemolyticus</i> SA2D2	WP_153645744.1
<i>V. parahaemolyticus</i> S042	WP_029837635
<i>V. alginolyticus</i> S6-61	WP_213874196
<i>V. cholerae</i> OY1_G01	WP_095466637
<i>V. alginolyticus</i> 2013V-1302	WP_154204368
<i>V. parahaemolyticus</i> S171	WP_025606966
<i>V. cholerae</i> YB2_A06	WP_057552372
<i>V. cholerae</i> VC22	WP_069503933
<i>V. cholerae</i> N2705-1	WP_148494788.1
<i>V. cholerae</i> N2705-2	WP_148494905.1
<i>V. cholerae</i> YB4_G06	WP_057552372
<i>V. diazotrophicus</i> 60.18M	WP_102986433



Supplementary Figure 2.2. Protein gels of purified proteins used in this study. (A) Wildtype NixI after purification with a heparin column, where eluate was used for the assays in this paper. (B) NixI^{N95A} purification with heparin column, where the eluate was used for the assays in this study. MW=molecular weight ladder (in kilodaltons), W1=wash 1, W2=wash 2, FT=flow through/unbound protein. Red boxes highlight the purified protein of interest.

A**B**

Supplementary Figure 2.3. NixI cannot cleave PCR products spanning cut regions from ICP1 (A) Native gels of PCR products incubated with decreasing concentrations of NixI with substrates spanning *gp121* and *gp156* cut regions which were used as the insert in the plasmids as in Figure 2C. (B) Alkaline denaturing agarose gel of the same reactions from A, with a nicked PCR product control and linear uncut control on the far left.



Supplementary Figure 2.4. NixI does not cleave PLE's rolling circle origin of replication, including PLEs putative packaging site (A) Gene map of PLE 1 showing relative position of PLEs RCR origin (ori). (B) Cleavage assay with a plasmid substrate containing 500bp stretching across PLE's RCR ori, with decreasing concentration of NixI labelled above the gel. MW is the molecular weight ladder.

Supplementary Table 3.1. Strains used in this study

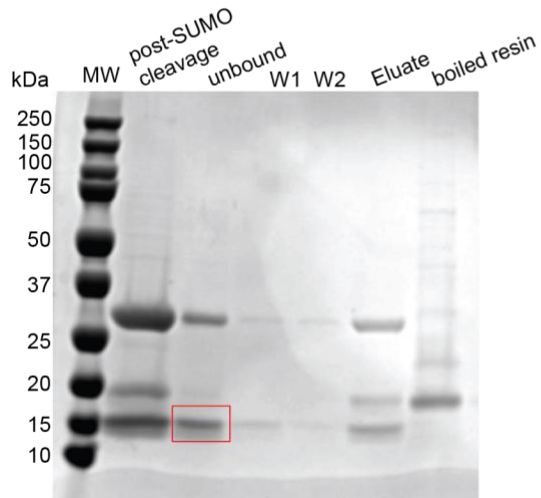
Strain Name	Description	Source
KDS6	<i>V. cholerae</i> E7946	Lab collection
KDS103	<i>V. cholerae</i> E7946 containing PLE1 Kan ^R	(32)
KS2421	<i>V. cholerae</i> E7946ΔlacZ::Spec ^R	(16)
KDS196	<i>V. cholerae</i> E7946 pKL06 empty vector Cm ^R	(100)
KL11	<i>V. cholerae</i> E7946 pKL06: P _{tac} nixI Cm ^R	(89)
KL20	<i>V. cholerae</i> E7946 pKL06: P _{tac} nixI+stiX Cm ^R	This study
KL124	BL21 pSUMO:His6-SUMO-stiX, Kan ^R	This study
KL806	BL21 pCold:stiX-His6, Kan ^R	This study
KL772-773	<i>V. cholerae</i> E7946 containing PLE1 Kan ^R + pKL06 empty vector, Cm ^R	This study
KDS163	<i>V. cholerae</i> E7946 containing PLE1ΔstiX Kan ^R	(100)
KL914-915	<i>V. cholerae</i> E7946 containing PLE1ΔstiX Kan ^R + pKL06 empty vector, Cm ^R	This study
KL916-917	<i>V. cholerae</i> E7946 containing PLE1ΔstiX Kan ^R + pKL06: P _{tac} stiX, Cm ^R	This study
KL735-736	<i>V. cholerae</i> E7946 containing PLE1+gp121 motifs Kan ^R	(89)
KL918-919	<i>V. cholerae</i> E7946 containing PLE1+gp121ΔstiX Kan ^R + pKL06 empty vector, Cm ^R	This study
KL920-921	<i>V. cholerae</i> E7946 containing PLE1+gp121ΔstiX Kan ^R + pKL06: P _{tac} stiX, Cm ^R	This study
KL907-908	<i>V. cholerae</i> E7946ΔclpP pKL06: P _{tac} nixI Cm ^R	This study
KL909-910	<i>V. cholerae</i> E7946ΔclpP pKL06: P _{tac} nixI+stiX Cm ^R	This study

Supplementary Table 3.2. Oligos used in this study

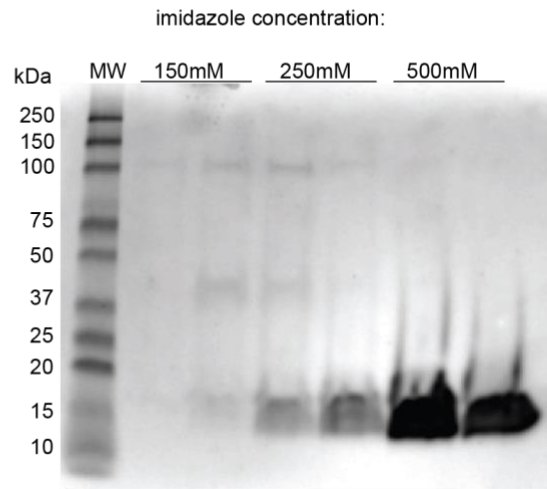
Primer	Sequence	Purpose	Source
zac14	AGGGTTTGAGTGCGATTACG	qPCR PLE	(32)
zac15	TGAGGTTTTACCACCTTTTGC	qPCR PLE	(32)
KS599	TGTTGGCATGTTTCGTATTTCC	NixI promoter EMSA probe	This study
Zac67	ATCAACCACCACCCACACC	NixI promoter EMSA probe	This study
KL241	GATGGAAGGGCTAACGGT	Gp121 EMSA probe	This study
KL242	CATCTCCGTGATTACCTTGGG	Gp121 EMSA probe	This study
KL348	CGCCACCATAATACTTTTGC AATGTC	Gp156 EMSA probe	(89)
KL349	GCCAGAGATTATGTCTGATG AAGAG	Gp156 EMSA probe	(89)
KL456	CACCATGAATCACAAAGTGA TGATAACCATCGAACTGTTA AACAG	Gibson primer to clone stiX onto pCold linearized with KL440/444	This study
KL457	GTGGTGATGATGGTGATGTT CTTTTGTTCCTTAAATTTGT TGTTGAGG	Gibson primer to clone stiX onto pCold linearized with KL440/444	This study
KL143	gaacagattggaggtATGATAACCAT CGAACTG	Gibson primer to clone stiX onto pSUMO linearized with ACM199/220	This study
KL144	gtggtggtggtggtgctcgaTCATTCTT TTGTTTCCTTAAATTTG	Gibson primer to clone stiX onto pSUMO linearized with ACM199/220	This study
KS951	CTGCTAAGGAGGTAACAACA AGATGCCAACACAAAAGACG TTAGACC	Gibson primer to clone nixI+stiX onto pKL06	(89)
KS966	ctctcatccgcaaaacagcTCATTCTT TTGTTTCCTTAAATTTGTTGT TG	Gibson primer to clone nixI+stiX onto pKL06	This study
KL55	AGGAGGTAACAACAAGATG ATAACCATCGAACTGTAAA C	Gibson primer to clone stiX onto pKL06, use with KS966	This study

A

StiX: post-Ni affinity purification, post-AKTA
SUMO protease cleaved

**B**

StiX-His6 Purification from pCold vector, elution fractions



Supplementary Figure 3.1. SDS-PAGE gels of purified StiX. A) StiX was affinity purified from a pSUMO vector with an N-terminus His6-SUMO cleavable tag. The SUMO tag was cleaved off and removed by re-binding the protein to Ni-NTA resin in batch purification. Unbound tag-free StiX (lane 3) was used for *in vitro* assays. B) StiX was re-purified from an inducible pCold vector with a C-terminus His6 tag (not cleavable). This fused tagged version of StiX was shown to be functional in abrogating NixI toxicity in *V. cholerae in vivo*. 500mM imidazole elution fractions were dialyzed and used in *in vitro* assays.

References

1. J. A. Fuhrman, Marine viruses and their biogeochemical and ecological effects. *Nature*. **399**, 541–548 (1999).
2. Ø. Bergh, K. Y. Børshiem, G. Bratbak, M. Heldal, High abundance of viruses found in aquatic environments. *Nature* (1989), doi:10.1038/340467a0.
3. S. van Houte, A. Buckling, E. R. Westra, Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiol. Mol. Biol. Rev.* **80**, 745–763 (2016).
4. K. D. Seed, S. M. Faruque, J. J. Mekalanos, S. B. Calderwood, F. Qadri, Phase Variable O Antigen Biosynthetic Genes Control Expression of the Major Protective Antigen and Bacteriophage Receptor in *Vibrio cholerae* O1. *PLoS Pathog.* **8**, e1002917 (2012).
5. K. D. Seed, M. Yen, B. Jesse Shapiro, I. J. Hilaire, R. C. Charles, J. E. Teng, L. C. Ivers, J. Boncy, J. B. Harris, A. Camilli, Evolutionary consequences of intra-patient phage predation on microbial populations. *Elife*. **3**, 1–10 (2014).
6. S. Van Houte, A. Buckling, E. R. Westra, Evolutionary Ecology of Prokaryotic Immune Mechanisms, doi:10.1128/MMBR.00011-16.
7. A. Bernheim, R. Sorek, The pan-immune system of bacteria: antiviral defence as a community resource. *Nat. Rev. Microbiol.* **18**, 113–119 (2020).
8. S. Doron, S. Melamed, G. Ofir, A. Leavitt, A. Lopatina, M. Keren, G. Amitai, R. Sorek, Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* (80-.). **359**, eaar4120 (2018).
9. E. V. Koonin, K. S. Makarova, Y. I. Wolf, M. Krupovic, Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. *Nat. Rev. Genet.* **21**, 119–131 (2020).
10. E. V. Koonin, K. S. Makarova, Y. I. Wolf, Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu. Rev. Microbiol.* **71**, 233–61 (2017).
11. L. Gao, H. Altae-Tran, F. Böhning, K. S. Makarova, M. Segel, J. L. Schmid-Burgk, J. Koob, Y. I. Wolf, E. V. Koonin, F. Zhang, Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* (80-.). **369**, 1077–1084 (2020).
12. A. Mutreja, D. W. Kim, N. R. Thomson, T. R. Connor, J. H. Lee, S. Kariuki, N. J. Croucher, S. Y. Choi, S. R. Harris, M. Lebens, S. Kumar Niyogi, E. J. Kim, T. Ramamurthy, J. Chun, J. L. N. Wood, J. D. Clemens, C. Czerkinsky, G. B. Nair, J. Holmgren, J. Parkhill, G. Dougan, Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature*. **477** (2011), doi:10.1038/nature10392.
13. M. Ali, A. R. Nelson, A. Lena Lopez, D. A. Sack, Updated Global Burden of Cholera in Endemic Countries. *PLoS Negl. Trop. Dis.* **9** (2015), doi:10.1371/journal.pntd.0003832.
14. K. D. Seed, K. L. Bodi, A. M. Kropinski, H. W. Ackermann, S. B. Calderwood, F. Qadri, A. Camilli, Evidence of a dominant lineage of *Vibrio cholerae*-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. *MBio*. **2**, e00334-10 (2011).
15. A. C. McKitterick, K. N. LeGault, A. Angermeyer, M. Alam, K. D. Seed, Competition between mobile genetic elements drives optimization of a phage-encoded CRISPR-Cas system: insights from a natural arms race. *Phil. Trans. R. Soc. B.* **374**, 20180089 (2019).
16. K. N. LeGault, S. G. Hays, A. Angermeyer, A. C. McKitterick, F. T. Johura, M. Sultana, T. Ahmed, M. Alam, K. D. Seed, Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts. *Science* (80-.). **373** (2021), doi:10.1126/science.abg2166.
17. C. M. Boyd, A. Angermeyer, S. G. Hays, Z. K. Barth, K. M. Patel, K. D. Seed, *Annu. Rev. Virol.*, in press, doi:10.1146/annurev-virology-091919-072020.

18. B. Lowey, A. T. Whiteley, A. F. A. Keszei, B. R. Morehouse, I. T. Mathews, S. P. Antine, V. J. Cabrera, D. Kashin, P. Niemann, M. Jain, F. Schwede, J. J. Mekalanos, S. Shao, A. S. Y. Lee, P. J. Kranzusch, CBASS Immunity Uses CARF-Related Effectors to Sense 3'-5'- and 2'-5'-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. *Cell* (2020), doi:10.1016/j.cell.2020.05.019.
19. A. Millman, A. Bernheim, A. Stokar-Avihail, T. Fedorenko, M. Voichek, A. Leavitt, Y. Oppenheimer-Shaanan, R. Sorek, Bacterial Retrons Function In Anti-Phage Defense. *Cell*. **183**, 1551-1561.e12 (2020).
20. J. E. Samson, A. H. Magadán, M. Sabri, S. Moineau, Revenge of the phages: Defeating bacterial defences. *Nat. Rev. Microbiol.* **11**, 675–687 (2013).
21. M. LeRoux, S. Srikant, M. L. Littlehale, G. Teodoro, S. Doron, M. Badiie, A. Leung, R. Sorek, M. T. Laub, *bioRxiv*, in press.
22. Y. Otsuka, T. Yonesaki, Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins. *Mol. Microbiol.* **83**, 669–681 (2012).
23. T. R. Blower, F. L. Short, P. C. Fineran, G. P. C. Salmond, Viral molecular mimicry circumvents abortive infection and suppresses bacterial suicide to make hosts permissive for replication. *Bacteriophage*. **2**, e23830 (2012).
24. M. D. Walkinshaw, P. Taylor, S. S. Sturrock, C. Atanasiu, T. Berge, R. M. Henderson, J. M. Edwardson, D. T. F. Dryden, Structure of Ocr from bacteriophage T7, a protein that mimics b-form DNA. *Mol. Cell*. **9**, 187–194 (2002).
25. A. Isaev, A. Drobiazko, N. Sierro, J. Gordeeva, I. Yosef, U. Qimron, N. V Ivanov, K. Severinov, Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence. *Nucleic Acids Res.* **48**, 5397–5406 (2020).
26. J. Bondy-Denomy, A. Pawluk, K. L. Maxwell, A. R. Davidson, Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*. **493** (2013), doi:10.1038/nature11723.
27. S. Gaba, D. Ebert, Time-shift experiments as a tool to study antagonistic coevolution. *Trends Ecol. Evol.* **24**, 226–232 (2009).
28. B. Koskella, M. A. Brockhurst, Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **38**, 916–931 (2014).
29. J. J. Dennehy, What Can Phages Tell Us about Host-Pathogen Coevolution? *Int. J. Evol. Biol.* **2012**, 1–12 (2012).
30. S. Paterson, T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, B. Libberton, A. Fenton, N. Hall, M. A. Brockhurst, Antagonistic coevolution accelerates molecular evolution. *Nature*. **464**, 275–278 (2010).
31. M. A. Brockhurst, T. Chapman, K. C. King, J. E. Mank, S. Paterson, G. D. D. Hurst, Running with the Red Queen: the role of biotic conflicts in evolution. *Proc. R. Soc. B Biol. Sci.* **281** (2014), doi:10.1098/rspb.2014.1382.
32. B. J. O'Hara, Z. K. Barth, A. C. McKitterick, K. D. Seed, A highly specific phage defense system is a conserved feature of the Vibrio cholerae mobilome. *PLoS Genet.* **13**, e1006838 (2017).
33. A. Angermeyer, S. G. Hays, M. H. T. Nguyen, F.-T. Johura, M. Sultana, M. Alam, K. D. Seed, Evolutionary sweeps of subviral parasites and their phage host bring unique parasite variants and disappearance of a phage CRISPR-Cas system. *bioRxiv*, 1–20 (2021).
34. K. D. Seed, D. W. Lazinski, S. B. Calderwood, A. Camilli, A bacteriophage encodes its

- own CRISPR/Cas adaptive response to evade host innate immunity. *Nature*. **494**, 489–491 (2013).
35. Z. Netter, C. M. Boyd, T. V. Silvas, K. D. Seed, A phage satellite tunes inducing phage gene expression using a domesticated endonuclease to balance inhibition and virion hijacking. *Nucleic Acids Res.* **49**, 4386–4401 (2021).
 36. S. W. Wilhelm, C. a Suttle, Viruses and Nutrient Cycles in the Sea aquatic food webs. *Bioscience*. **49**, 781–788 (1999).
 37. B. Koskella, New approaches to characterizing bacteria–phage interactions in microbial communities and microbiomes. *Environ. Microbiol. Rep.* **11**, 15–16 (2019).
 38. F. Rodriguez-Valera, A. B. Martin-Cuadrado, B. Rodriguez-Brito, L. Pašić, T. F. Thingstad, F. Rohwer, A. Mira, Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* **7**, 828–836 (2009).
 39. Y. Zheng, R. J. Roberts, S. Kasif, Identification of genes with fast-evolving regions in microbial genomes. *Nucleic Acids Res.* **32**, 6347–6357 (2004).
 40. G. Ofir, S. Melamed, H. Sberro, Z. Mukamel, S. Silverman, G. Yaakov, S. Doron, R. Sorek, DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nat. Microbiol.* (2017), doi:10.1038/s41564-017-0051-0.
 41. T. Goldfarb, H. Sberro, E. Weinstock, O. Cohen, S. Doron, Y. Charpak-Amikam, S. Afik, G. Ofir, R. Sorek, BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* **34**, 169–183 (2015).
 42. S. M. Faruque, M. J. Islam, Q. S. Ahmad, A. S. G. Faruque, D. A. Sack, G. B. Nair, J. J. Mekalanos, Self-limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. *Proc. Natl. Acad. Sci.* **102**, 6119–6124 (2005).
 43. M. A. Jensen, S. M. Faruque, J. J. Mekalanos, B. R. Levin, Modeling the role of bacteriophage in the control of cholera outbreaks. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4652–4657 (2006).
 44. Z. K. Barth, M. H. T. Nguyen, K. D. Seed, A chimeric nuclease substitutes CRISPR-Cas : A phage weaponizes laterally acquired specificity to destroy subviral parasites. *bioRxiv* (2021), doi:https://doi.org/10.1101/2021.02.21.432181.
 45. M. K. Waldor, H. Tschäpe, J. J. Mekalanos, A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. *J. Bacteriol.* **178**, 4157–4165 (1996).
 46. R. A. F. Wozniak, D. E. Fouts, M. Spagnoletti, M. M. Colombo, D. Ceccarelli, Comparative ICE Genomics: Insights into the Evolution of the SXT/R391 Family of ICEs. *PLoS Genet.* **5**, 1000786 (2009).
 47. X. Li, Y. Du, P. Du, H. Dai, Y. Fang, Z. Li, N. Lv, B. Zhu, B. Kan, D. Wang, SXT/R391 integrative and conjugative elements in *Proteus* species reveal abundant genetic diversity and multidrug resistance. *Nat. Publ. Gr.* (2016), doi:10.1038/srep37372.
 48. V. Burrus, J. Marrero, M. K. Waldor, The current ICE age: Biology and evolution of SXT-related integrating conjugative elements. *Plasmid.* **55**, 173–183 (2006).
 49. J. Gordeeva, N. Morozova, N. Sierro, A. Isaev, T. Sinkunas, K. Tsvetkova, M. Matlashov, L. Truncaite, R. D. Morgan, N. V. Ivanov, V. Siksnys, L. Zeng, K. Severinov, BREX system of *Escherichia coli* distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Res.* **47**, 253–265 (2019).
 50. C. Tai, J. Li, H. Ou, X. Li, D. Bi, M. Liu, J. Sun, Y. Xie, Z. Deng, ICEberg 2.0: An updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* **47**

- (2019), doi:10.1093/nar/gky1123.
51. A. Bioteau, R. Durand, V. Burrus, Redefinition and Unification of the SXT/R391 Family of Integrative and Conjugative Elements. *Appl. Environ. Microbiol.* **84**, 1–16 (2018).
 52. S. Slattery, J. Tony Pembroke, J. G. Murnane, M. P. Ryan, isolation, nucleotide sequencing and genomic comparison of a Novel SXT/R391 ICE mobile genetic element isolated from a municipal wastewater environment. *Sci. Rep.* (2020), doi:10.1038/s41598-020-65216-5.
 53. K. S. Makarova, V. Anantharaman, N. V. Grishin, E. V. Koonin, L. Aravind, CARF and WYL domains: Ligand-binding regulators of prokaryotic defense systems. *Front. Genet.* **5**, 1–9 (2014).
 54. M. Yen, L. S. Cairns, A. Camilli, A cocktail of three virulent bacteriophages prevents *Vibrio cholerae* infection in animal models. *Nat. Commun.* **8**, 14187 (2017).
 55. L. D. R. Melo, H. Oliveira, D. P. Pires, K. Dabrowska, J. Azeredo, Phage therapy efficacy: a review of the last 10 years of preclinical studies. *Crit. Rev. Microbiol.* **46**, 78–99 (2020).
 56. L. D. R. Melo, P. Veiga, N. Cerca, A. M. Kropinski, C. Almeida, J. Azeredo, S. Sillankorva, Development of a phage cocktail to control *Proteus mirabilis* catheter-associated urinary tract infections. *Front. Microbiol.* **7**, 1–12 (2016).
 57. M. Spagnoletti, D. Ceccarelli, A. Rieux, M. Fondi, E. Taviani, R. Fani, M. M. Colombo, R. R. Colwell, F. Balloux, Acquisition and Evolution of SXT-R391 Integrative Conjugative Elements in the Seventh-Pandemic *Vibrio cholerae* Lineage. *MBio.* **5** (2014), doi:10.1128/mBio.01356-14.
 58. R. Wang, D. Yu, J. Yue, B. Kan, Variations in SXT elements in epidemic *Vibrio cholerae* O1 El Tor strains in China. *Sci. Rep.* **6**, 1–8 (2016).
 59. Z. K. Barth, Z. Netter, A. Angermeyer, P. Bhardwaj, K. D. Seed, A family of viral satellites manipulates invading virus gene expression and affects cholera toxin mobilization. *mSystems.* **5**, e00358-20 (2020).
 60. M. K. Waldor, J. J. Mekalanos, Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science.* **272**, 1910–4 (1996).
 61. A. Fomenkov, K. D. Lunnen, Z. Zhu, B. P. Anton, G. G. Wilson, T. Vincze, R. J. Roberts, Complete Genome Sequence and Methyloome Analysis of *Bacillus* Strain X1. *Genome Announc.* **3** (2015), doi:10.1128/genomeA.01593-14.
 62. G. Garriss, M. K. Waldor, V. Burrus, Mobile Antibiotic Resistance Encoding Elements Promote Their Own Diversity. *PLoS Genet.* **5**, 1000775 (2009).
 63. A. L. Rodríguez-Blanco Manuel Lemos Carlos R Osorio, Unveiling the pan-genome of the SXT/R391 family of ICEs: molecular characterisation of new variable regions of SXT/R391-like ICEs detected in *Pseudoalteromonas* sp. and *Vibrio scophthalmi*. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **109**, 1141–1152 (2016).
 64. M. Balado, M. L. Lemos, C. R. Osorio, Integrating conjugative elements of the SXT/R391 family from fish-isolated *Vibrios* encode restriction-modification systems that confer resistance to bacteriophages. *FEMS Microbiol. Ecol.* **83**, 457–467 (2013).
 65. V. Burrus, M. K. Waldor, Control of SXT Integration and Excision. *J. Bacteriol.* **185**, 5045–5054 (2003).
 66. J. W. Beaber, B. Hochhut, M. K. Waldor, SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature.* **427**, 72–74 (2004).
 67. L. Alexandrova, F. Haque, P. Rodriguez, A. C. Marrazzo, J. A. Grembi, V.

- Ramachandran, A. J. Hryckowian, C. M. Adams, M. S. A. Siddique, A. I. Khan, F. Qadri, J. R. Andrews, M. Rahman, A. M. Spormann, G. K. Schoolnik, A. Chien, E. J. Nelson, Identification of Widespread Antibiotic Exposure in Patients with Cholera Correlates with Clinically Relevant Microbiota Changes. *J. Infect. Dis.* **220**, 1655–1666 (2019).
68. R. Baddam, N. Sarker, D. Ahmed, R. Mazumder, A. Abdullah, R. Morshed, A. Hussain, S. Begum, L. Shahrin, A. I. Khan, M. S. Islam, T. Ahmed, M. Alam, J. D. Clemens, N. Ahmed, Genome Dynamics of *Vibrio cholerae* Isolates Linked to Seasonal Outbreaks of Cholera in Dhaka, Bangladesh. *MBio.* **11**, 1–14 (2020).
 69. A. Sarkar, D. Morita, A. Ghosh, G. Chowdhury, A. K. Mukhopadhyay, K. Okamoto, T. Ramamurthy, Altered Integrative and Conjugative Elements (ICEs) in Recent *Vibrio cholerae* O1 Isolated From Cholera Cases, Kolkata, India. *Front. Microbiol.* **10** (2019), doi:10.3389/fmicb.2019.02072.
 70. A. B. Dalia, K. D. Seed, S. B. Calderwood, A. Camilli, A globally distributed mobile genetic element inhibits natural transformation of *Vibrio cholerae*. *Proc. Natl. Acad. Sci.* **112** (2015), doi:10.1073/pnas.1509097112.
 71. E. Bordeleau, E. Brouillette, N. Robichaud, V. Burrus, Beyond antibiotic resistance: Integrating conjugative elements of the SXT/R391 family that encode novel diguanylate cyclases participate to c-di-gmp signalling in *Vibrio cholerae*. *Environ. Microbiol.* **12**, 510–523 (2010).
 72. M. P. Ryan, S. Slattey, J. T. Pembroke, A novel arsenate-resistant determinant associated with ICEpMERPH, a member of the SXT/R391 group of mobile genetic elements. *Genes (Basel)*. **10**, 1–12 (2019).
 73. S. E. Peters, J. L. Hobman, P. Strike, D. A. Ritchie, Novel mercury resistance determinants carried by IncJ plasmids pMERPH and R391. *MGG Mol. Gen. Genet.* **228**, 294–299 (1991).
 74. A. Betts, O. Kaltz, M. E. Hochberg, Contrasted coevolutionary dynamics between a bacterial pathogen and its bacteriophages. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15279 (2014).
 75. A. Papkou, T. Guzella, W. Yang, S. Koepper, B. Pees, R. Schalkowski, The genomic basis of Red Queen dynamics during rapid reciprocal host – pathogen coevolution. **116**, 923–928 (2019).
 76. J. P. J. Hall, E. Harrison, K. Pärnänen, M. Virta, M. A. Brockhurst, The Impact of Mercury Selection and Conjugative Genetic Elements on Community Structure and Resistance Gene Transfer. *Front. Microbiol.* **11**, 1–14 (2020).
 77. P. H. Oliveira, M. Touchon, E. P. C. Rocha, Regulation of genetic flux between bacteria by restriction-modification systems. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 5658–5663 (2016).
 78. J. B. Kaper, J. G. Morris, M. M. Levine, Cholera. *Clin. Microbiol. Rev.* **8**, 48–86 (1995).
 79. S. M. Faruque, I. Bin Naser, M. J. Islam, A. S. G. Faruque, A. N. Ghosh, G. Balakrish Nair, D. A. Sack, J. J. Mekalanos, Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. *Proc. Natl. Acad. Sci.* **102** (2005).
 80. S. Takano, K. Fukuda, A. Koto, R. Miyazaki, A novel system of bacterial cell division arrest implicated in horizontal transmission of an integrative and conjugative element. *PLoS Genet.* **15** (2019), doi:10.1371/journal.pgen.1008445.
 81. M. Alam, M. Sultana, G. B. Nair, R. B. Sack, D. A. Sack, A. K. Siddique, A. Ali, A. Huq, R. R. Colwell, Toxigenic *Vibrio cholerae* in the aquatic environment of Mathbaria,

- Bangladesh. *Appl. Environ. Microbiol.* **72**, 2849–2855 (2006).
82. A. C. McKitterick, K. D. Seed, Anti-phage islands force their target phage to directly mediate island excision and spread. *Nat. Commun.* **9**, e2348 (2018).
 83. A. B. Dalia, D. W. Lazinski, A. Camilli, Identification of a Membrane-Bound Transcriptional Regulator That Links Chitin and Natural Competence in *Vibrio cholerae*. *MBio.* **5**, e01028-13 (2014).
 84. A. M. Box, M. J. McGuffie, B. J. O'Hara, K. D. Seed, Functional Analysis of Bacteriophage Immunity through a Type I-E CRISPR-Cas System in *Vibrio cholerae* and Its Application in Bacteriophage Genome Engineering. *J. Bacteriol.* **198** (2016), doi:10.1128/JB.00747-15.
 85. A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, P. A. Pevzner, SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
 86. S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. E. Tosatto, R. D. Finn, The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
 87. R. J. Roberts, T. Vincze, J. Posfai, D. Macelis, REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res.* **43**, D298–D299 (2015).
 88. A. Roy, A. Kucukural, Y. Zhang, I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
 89. K. N. LeGault, Z. K. Barth, P. DePaola, K. D. Seed, A phage parasite deploys a nicking nuclease effector to inhibit viral host replication. *Nucleic Acids Res.*, 1–17 (2022).
 90. J. Chen, R. P. Novick, Phage-mediated intergeneric transfer of toxin genes. *Science* (80-). **323**, 139–141 (2009).
 91. A. Fillol-Salom, L. Miguel-Romero, A. Marina, J. Chen, J. R. Penadés, Beyond the CRISPR-Cas safeguard: PICI-encoded innate immune systems protect bacteria from bacteriophage predation. *Curr. Opin. Microbiol.* **56** (2020), pp. 52–58.
 92. M. A. Tormo-Más, I. Mir, A. Shrestha, S. M. Tallent, S. Campoy, I. Lasa, J. Barbé, R. P. Novick, G. E. Christie, J. R. Penadés, Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature.* **465**, 779–782 (2010).
 93. C. Úbeda, P. Barry, J. R. Penadés, R. P. Novick, A pathogenicity island replicon in *Staphylococcus aureus* replicates as an unstable plasmid. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14182–14188 (2007).
 94. B. H. Lindqvist, G. Deho, R. Calendar, Mechanisms of Genome Propagation and Helper Exploitation by Satellite Phage P4. *Microbiol. Rev.* **57**, 683–702 (1993).
 95. S. M. Tallent, T. B. Langston, R. G. Moran, G. E. Christie, Transducing Particles of *Staphylococcus aureus* Pathogenicity Island SaPI1 Are Comprised of Helper Phage-Encoded Proteins. *J. Bacteriol.* **189**, 7520–7524 (2007).
 96. J. R. Penadés, G. E. Christie, The Phage-Inducible Chromosomal Islands: A Family of Highly Evolved Molecular Parasites. *Annu. Rev. Virol.* **2** (2015), doi:10.1146/annurev-virology-031413-085446.
 97. J. A. Lindsay, A. Ruzin, H. F. Ross, N. Kurepina, R. P. Novick, The gene for toxic shock

- toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol. Microbiol.* **29**, 527–43 (1998).
98. A. Fillol-Salom, R. Martínez-Rubio, R. F. Abdulrahman, J. Chen, R. Davies, J. R. Penadés, Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. *ISME J.* **12**, 2114–2128 (2018).
 99. T. Hackl, R. Laurenceau, M. J. Ankenbrand, C. Bliem, Z. Cariani, E. Thomas, K. D. Dooley, A. A. Arellano, S. L. Hogle, P. Berube, G. E. Leventhal, E. Luo, J. Eppley, A. A. Zayed, J. Beaulaurier, R. Stepanauskas, M. B. Sullivan, E. F. DeLong, S. J. Biller, S. W. Chisholm, *bioRxiv*, in press.
 100. S. G. Hays, K. D. Seed, Dominant *Vibrio cholerae* phage exhibits lysis inhibition sensitive to disruption by a defensive phage satellite. *Elife.* **9**, e53200 (2020).
 101. Z. K. Barth, T. V. Silvas, A. Angermeyer, K. D. Seed, Genome replication dynamics of a bacteriophage and its satellite reveal strategies for parasitism and viral restriction. *Nucleic Acids Res.* **48**, 249–263 (2019).
 102. Z. K. Barth, M. H. T. Nguyen, K. D. Seed, A chimeric nuclease substitutes a phage CRISPR-Cas system to provide sequence- specific immunity against subviral parasites. *Elife.* **10**, e68339 (2021).
 103. A. B. Hickman, D. R. Ronning, R. M. Kotin, F. Dyda, Structural unity among viral origin binding proteins: Crystal structure of the nuclease domain of adeno-associated virus Rep. *Mol. Cell.* **10**, 327–337 (2002).
 104. S. A. Khan, Plasmid rolling-circle replication: Highlights of two decades of research. *Plasmid.* **53**, 126–136 (2005).
 105. M. Narajczyk, S. Barajska, A. Wdgrzyn, G. Wdgrzyn, Switch from to replication of bacteriophage DNA: factors involved in the process and a model for its regulation. *Mol Genet Genomics.* **278**, 65–74 (2007).
 106. V. B. Rao, M. Feiss, Mechanisms of DNA Packaging by Large Double-Stranded DNA Viruses. *Annu. Rev. Virol.* **2**, 351–78 (2015).
 107. P. H. Culviner, M. T. Laub, Global Analysis of the *E. coli* Toxin MazF Reveals Widespread Cleavage of mRNA and the Inhibition of rRNA Maturation and Ribosome Biogenesis. *Mol. Cell.* **70**, 868–880 (2018).
 108. B. W. Shen, M. Landthaler, D. A. Shub, B. L. Stoddard, DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.* **342**, 43–56 (2004).
 109. M. Landthaler, B. W. Shen, B. L. Stoddard, D. A. Shub, I-BasI and I-HmuI: Two Phage Intron-encoded Endonucleases with Homologous DNA Recognition Sequences but Distinct DNA Specificities. *J. Mol. Biol.* **358**, 1137–1151 (2006).
 110. M. Remmert, A. Biegert, A. Hauser, J. Söding, HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods.* **9**, 173–175 (2012).
 111. P. K. Damle, E. A. Wall, M. S. Spilman, A. D. Dearborn, G. Ram, R. P. Novick, T. Dokland, G. E. Christie, The roles of SaPI1 proteins gp7 (CpmA) and gp6 (CpmB) in capsid size determination and helper phage interference. *Virology.* **432**, 277–282 (2012).
 112. R. P. Novick, G. E. Christie, J. R. Penadés, The phage-related chromosomal islands of Gram-positive bacteria. *Nat. Rev. Microbiol.* **8**, 541–551 (2010).
 113. R. C. Matos, N. Lapaque, L. Rigottier-Gois, L. Debarbieux, T. Meylheuc, B. Gonzalez-Zorn, F. Repoila, M. De, F. Lopes, P. Serror, *Enterococcus faecalis* Prophage Dynamics and Contributions to Pathogenic Traits. *PLoS Genet.* **9**, e1003539 (2013).
 114. R. Cheng, F. Huang, H. Wu, X. Lu, Y. Yan, B. Yu, X. Wang, B. Zhu, A nucleotide-

- sensing endonuclease from the Gabija bacterial defense system. *Nucleic Acids Res.* **49**, 5216–5229 (2021).
115. X. Xiong, G. Wu, Y. Wei, L. Liu, Y. Zhang, R. Su, X. Jiang, M. Li, H. Gao, X. Tian, Y. Zhang, L. Hu, S. Chen, Y. Tang, S. Jiang, R. Huang, Z. Li, Y. Wang, Z. Deng, J. Wang, P. C. Dedon, S. Chen, L. Wang, SspABCD-SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities. *Nat. Microbiol.* **5**, 917–928 (2020).
 116. K. B. Vrtis, J. M. Dewar, G. Chistol, R. A. Wu, T. G. W. Graham, J. C. Walter, Single-strand DNA breaks cause replisome disassembly. *Mol. Cell.* **81**, 1309–1318 (2021).
 117. M. M. Cox, Recombinational DNA repair in bacteria and the RecA protein. *Prog. Nucleic Acid Res. Mol. Biol.* **63**, 311–366 (1999).
 118. A. C. McKitterick, S. G. Hays, F. T. Johura, M. Alam, K. D. Seed, Viral Satellites Exploit Phage Proteins to Escape Degradation of the Bacterial Host Chromosome. *Cell Host Microbe.* **26**, 504–514 (2019).
 119. N. Carpena, K. A. Manning, T. Dokland, A. Marina, J. R. Penadés, Convergent evolution of pathogenicity islands in helper cos phage interference. *Phil. Trans. R. Soc. B.* **371**, 20150505 (2016).
 120. R. Martínez-Rubio, N. Quiles-Puchalt, M. Martí, S. Humphrey, G. Ram, D. Smyth, J. Chen, R. P. Novick, J. R. Penadés, Phage-inducible islands in the Gram-positive cocci. *ISME J.* **11**, 1029–1042 (2017).
 121. C. Weigel, H. Seitz, Bacteriophage replication modules. *FEMS Microbiol. Rev.* **30**, 321–381 (2006).
 122. G. Ram, J. Chen, K. Kumar, H. F. Ross, C. Ubeda, P. K. Damle, K. D. Lane, J. R. Penadés, G. E. Christie, R. P. Novick, Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc. Natl. Acad. Sci.* **109**, 16300–16305 (2012).
 123. A. Fillol-Salom, J. Bacarizo, M. Alqasmi, J. R. Ciges-Tomas, R. Martínez-Rubio, A. W. Roszak, R. J. Cogdell, J. Chen, A. Marina, J. R. Penadés, Hijacking the Hijackers: Escherichia coli Pathogenicity Islands Redirect Helper Phage Packaging for Their Own Benefit. *Mol. Cell.* **75**, 1020–1030 (2019).
 124. K. J. Forsberg, D. T. Schmidtke, R. Werther, R. V. Uribe, D. Hausman, M. O. A. Sommer, B. L. Stoddard, B. K. Kaiser, H. S. Malik, *The novel anti-CRISPR AcrIIA22 relieves DNA torsion in target plasmids and impairs SpyCas9 activity* (2021), vol. 19.
 125. J. A. Moura de Sousa, E. P. C. Rocha, To catch a hijacker: abundance, evolution and genetic diversity of P4-like bacteriophage satellites. *bioRxiv* (2021), doi:<https://doi.org/10.1101/2021.03.30.437493>.
 126. F. Rousset, J. Dowding, A. Bernheim, E. P. C. Rocha, D. Bikard, *bioRxiv*, in press, doi:10.1101/2021.01.21.427644.
 127. K. M. Kauffman, J. M. Brown, R. S. Sharma, D. Vaninsberghe, J. Elsherbini, M. Polz, L. Kelly, Viruses of the Nahant Collection, characterization of 251 marine Vibrionaceae viruses. *Sci. Data.* **5**, 180114 (2018).
 128. L. Mateus, L. Costa, Y. J. Silva, C. Pereira, A. Cunha, A. Almeida, Efficiency of phage cocktails in the inactivation of Vibrio in aquaculture. *Aquaculture.* **424–425**, 167–173 (2014).
 129. E. V. Koonin, Molecular Cell Spotlight Hunting for Treasure Chests in Microbial Defense Islands (2018), doi:10.1016/j.molcel.2018.05.025.
 130. A. L. Bryson, Y. Hwang, S. Sherrill-Mix, G. D. Wu, J. D. Lewis, L. Black, T. A. Clark, F.

- D. Bushman, Covalent modification of bacteriophage T4 DNA inhibits CRISPRCas9. *MBio*. **6**, 1–9 (2015).
131. V. Chaikereetisak, K. Nguyen, K. Khanna, A. F. Brilot, M. L. Erb, J. K. C. Coker, A. Vavilina, G. L. Newton, R. Buschauer, K. Pogliano, E. Villa, D. A. Agard, J. Pogliano, Assembly of a nucleus-like structure during viral replication in bacteria. *Science (80-.)*. **355**, 194–197 (2017).
 132. M. R. Tock, D. T. F. Dryden, The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* **8**, 466–472 (2005).
 133. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* (2021), doi:10.1038/s41586-021-03819-2.
 134. L. Aravind, V. Anantharaman, S. Balaji, M. M. Babu, L. M. Iyer, The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiol. Rev.* **29**, 231–262 (2005).
 135. L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
 136. C. Fang, L. Li, L. Shen, J. Shi, S. Wang, Y. Feng, Y. Zhang, Structures and mechanism of transcription initiation by bacterial ECF factors. *Nucleic Acids Res.* **47**, 7094–7104 (2019).
 137. A. Harms, D. E. Brodersen, N. Mitarai, K. Gerdes, Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Mol. Cell.* **70**, 768–784 (2018).
 138. J. Kim, F. M. Raushel, Allosteric control of the oligomerization of carbamoyl phosphate synthetase from *Escherichia coli*. *Biochemistry.* **40**, 11030–11036 (2001).
 139. C. D. Aakre, T. N. Phung, D. Huang, M. T. Laub, A Bacterial Toxin Inhibits DNA Replication Elongation through a Direct Interaction with the β Sliding Clamp. *Mol. Cell.* **52**, 617–628 (2013).
 140. A. F. Haag, M. Podkowik, R. Ibarra-Chávez, F. Gallego del Sol, G. Ram, J. Chen, A. Marina, R. P. Novick, J. R. Penadés, A regulatory cascade controls *Staphylococcus aureus* pathogenicity island activation. *Nat. Microbiol.* **2021**, 1–9 (2021).
 141. B. Roucourt, R. Lavigne, The role of interactions between phage and bacterial proteins within the infected cell: A diverse and puzzling interactome. *Environ. Microbiol.* **11**, 2789–2805 (2009).
 142. C. F. Schuster, R. Bertram, Toxin–antitoxin systems are ubiquitous and versatile modulators of prokaryotic cell fate, doi:10.1111/1574-6968.12074.
 143. A. M. Hall, B. Gollan, S. Helaine, Toxin–antitoxin systems: reversible toxicity. *Curr. Opin. Microbiol.* **36**, 102–110 (2017).
 144. M. LeRoux, P. H. Culviner, Y. J. Liu, M. L. Littlehale, M. T. Laub, Stress Can Induce Transcription of Toxin-Antitoxin Systems without Activating Toxin. *Mol. Cell.* **79**, 280–292.e8 (2020).
 145. S. Ronneau, S. Helaine, Clarifying the Link between Toxin–Antitoxin Modules and Bacterial Persistence. *J. Mol. Biol.* **431**, 3462–3471 (2019).
 146. M. Leroux, M. T. Laub, Toxin-Antitoxin Systems as Phage Defense Elements, 21–43

- (2022).
147. S. C. Smith, K. K. Joshi, J. J. Zik, K. Trinh, A. Kamajaya, P. Chien, K. R. Ryan, Cell cycle-dependent adaptor complex for ClpXP-mediated proteolysis directly integrates phosphorylation and second messenger signals. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14229–14234 (2014).
 148. E. L. Flannery, S. M. Antczak, H. L. T. Mobley, Self-Transmissibility of the Integrative and Conjugative Element ICEPm1 between Clinical Isolates Requires a Functional Integrase, Relaxase, and Type IV Secretion System. *J. Bacteriol.* **193**, 4104–4112 (2011).
 149. L. Bie, H. Wu, X. H. Wang, M. Wang, H. Xu, Identification and characterization of new members of the SXT/R391 family of integrative and conjugative elements (ICEs) in *Proteus mirabilis*. *Int. J. Antimicrob. Agents.* **50**, 242–246 (2017).
 150. B. R. Morehouse, A. A. Govande, A. Millman, A. F. A. Keszei, B. Lowey, G. Ofir, S. Shao, R. Sorek, P. J. Kranzusch, STING cyclic dinucleotide sensing originated in bacteria. *Nature* (2020), doi:10.1038/s41586-020-2719-5.
 151. A. T. Whiteley, J. B. Eaglesham, C. C. De Oliveira Mann, B. R. Morehouse, B. Lowey, E. A. Nieminen, O. Danilchanka, D. S. King, A. S. Y. Lee, J. J. Mekalanos, P. J. Kranzusch, Bacterial cGAS-like enzymes synthesize diverse nucleotide signals Discovery of a pyrimidine-containing CDN, doi:10.1038/s41586-019-0953-5.
 152. A. G. Johnson, T. Wein, M. L. Mayer, B. Duncan-Lowey, E. Yirmiya, Y. Oppenheimer-Shaanan, G. Amitai, R. Sorek, P. J. Kranzusch, Bacterial gasdermins reveal an ancient mechanism of cell death. *Science (80-.)*. **375**, 221–225 (2022).
 153. T. Wein, R. Sorek, Bacterial origins of human cell-autonomous innate immune mechanisms, doi:10.1038/s41577-022-00705-4.
 154. J. Bobonis, K. Mitosch, A. Mateus, G. Kritikos, J. R. Elfenbein, M. M. Savitski, H. Andrews-Polymenis, A. Typas, *bioRxiv*, in press.
 155. C. K. Guegler, M. T. Laub, Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection. *Mol. Cell.* **81**, 2361-2373.e9 (2021).
 156. S. Srikant, C. K. Guegler, M. T. Laub, The evolution of a counter-defense mechanism in a virus constrains its host range Keywords (2022).
 157. A. Bioteau, R. Durand, crossm Integrative and Conjugative Elements. **84**, 1–16 (2018).