

Small protein modules dictate prophage fates during polylysogeny

Justin E. Silpe^{1,2}, Olivia P. Duddy¹, Grace E. Johnson^{1,2}, Grace A. Beggs¹, Fatima A. Hussain³, Kevin J. Forsberg⁴, and Bonnie L. Bassler^{1,2,#}

1 Department of Molecular Biology, Princeton University, Princeton, NJ, USA

2 Howard Hughes Medical Institute, Chevy Chase, MD, USA

3 Massachusetts Institute of Technology, Cambridge, MA, USA

4 University of Texas Southwestern Medical Center, Dallas, TX, USA

Correspondence: bbassler@princeton.edu

SUMMARY

Most bacteria in the biosphere are predicted to be polylysogens, harboring multiple prophages. In studied systems, prophage induction from lysogeny to lysis is near-universally driven by DNA-damaging agents. Thus, how co-residing prophages compete for cell resources if they respond to an identical trigger is unknown. Here, we discover regulatory modules that control prophage induction independently of the DNA damage cue. The modules bear little resemblance at the sequence level but share a regulatory logic by having a transcription factor that activates the expression of a neighboring gene encoding a small protein. The small protein inactivates the master repressor of lysis, leading to induction. Polylysogens harboring two prophages exposed to DNA damage release mixed populations of phages. Single-cell analyses reveal that this blend is a consequence of discrete subsets of cells producing one, the other, or both phages. By contrast, induction via the DNA-damage-independent module results in cells producing only the phage sensitive to that specific cue. Thus, in the polylysogens tested, the stimulus used to induce lysis determines phage productivity. Considering the lack of potent DNA-damaging agents in natural habitats, additional phage-encoded sensory pathways to lysis likely play fundamental roles in phage-host biology and inter-prophage competition.

INTRODUCTION

Phages are viruses that infect bacteria and are important drivers of bacterial diversity and microbial community development.^{1,2} Following host infection, temperate phages can undertake one of two lifestyles. They can enter the lytic cycle, in which they exploit host resources, replicate, produce viral particles, and kill their host.^{3,4} Alternatively, phages can enter the lysogenic state and remain dormant (i.e., as prophages), and they are passed down to host progeny.⁵ Phages can also cause chronic host infections in which they persist and extrude viral particles without killing the host.^{6,7}

During lysogeny, model temperate phages, including phage lambda, produce a repressor (called *ci*) that binds to and prevents expression from a promoter (termed P_R) that controls the lysis genes.⁴ Following DNA damage, activation of the host RecA protein leads to autoproteolysis and inactivation of the *ci* repressor.^{8,9} Consequently, P_R is de-repressed, triggering phage replication, host-cell lysis, and transmission of the phage to neighboring bacteria. Other temperate phages harbor repressors lacking the peptidase domain responsible for autoproteolysis. Rather, a peptidase or antirepressor encoded elsewhere in the phage genome is activated by the host SOS response.^{10,11} The understanding that all bacteria possess *recA*, coupled with the fact that phages are omnipresent, has led to the common view that the host SOS response is the 'universal' prophage inducer. However, significant concentrations of potent DNA-damaging agents are rare in the environment, and increasingly, phages are being discovered that are not induced by DNA damage.¹² Together, these findings suggest that undiscovered induction triggers exist in nature.

Recent findings reveal that quorum-sensing (QS) signals represent one SOS-independent induction trigger for phage lysis-lysogeny lifestyle transitions.^{13–15} QS is a process of cell-to-cell communication that bacteria use to orchestrate collective behaviors. QS relies on the production,

release, and group-wide detection of and response to extracellular signaling molecules called autoinducers (AI).¹⁶ Phages can harbor phage-to-phage QS-like communication systems, such as the arbitrium system in SP β phages,¹⁵ or as in vibriophage VP882, they can monitor host bacterial QS-mediated communication pathways to tune the timing of the lysogeny to lysis switch to changes in host-cell density.¹⁴ Phage VP882 is a linear plasmid-like prophage that encodes a homolog of the *Vibrio* QS receptor VqmA, called VqmA_{Phage}, which is activated by a host-produced AI, called DPO.^{14,17} Upon binding DPO, VqmA_{Phage} activates transcription of a counter-oriented gene called *qtip*. Production of Qtip launches host-cell lysis. Our hypothesis is that, by surveilling a bacterial-produced QS signal, the phage can integrate host-cell density information into its decision-making process, and by lysing its host at high cell density, the phage maximizes the probability of infecting other cells in the population.

Bacteria commonly harbor multiple prophages, a state called polylysogeny. How prophages residing in a polylysogenic host compete for host resources is not well understood. What is known is that following DNA damage, the number of virions released for lambdoid prophages is lower in polylysogens than in monolysogens.¹⁸ This result suggests that in polylysogens, co-induced prophages compete for reproductive success. It is possible that in the context of polylysogeny, if a prophage possesses an alternative, non-DNA damage-dependent pathway to lysis, it could compete more effectively with co-residing prophages that cannot respond to the alternative induction cue. Here, we sought to identify and characterize polylysogenic bacteria harboring prophages possessing SOS-dependent and SOS-independent pathways to lysis. Our aim was to use them as models to explore within-host prophage competition.

The present work describes the discovery of SOS-independent, phage-encoded lysis-inducing modules which, despite bearing little resemblance to one another at the sequence level, share a common regulatory logic. Our characterization of several of these phages shows they all employ

a transcription factor to activate the expression of a divergently transcribed gene encoding a small protein (smORF). The smORF launches the transition from lysogeny to lysis. The smORFs studied here lack homologs and predictable domains, and yet they operate by inactivating the same respective target, the *ci* repressor in the phage that encodes the smORF. The mechanisms by which the transcription factors regulate their partner *smORF* genes vary. In some cases, the transcription factors operate independently. In other cases, *smORF* gene expression requires a xenobiotic responsive element (XRE) family protein working in conjunction with a LuxR-type QS receptor/transcription factor that requires a bacterial-produced AI ligand for activity.

The prophage-containing isolates on which we focus are polylysogenic. We show that the addition of a DNA-damaging agent leads to phage-mediated lysis of the bacteria and release of a mixed population of phage particles. Single cell analyses demonstrate that this outcome stems from a subset of host cells expressing lytic genes from only one of the phages, another subset expressing lytic genes from only the other phage, and the final subset of cells expressing lytic genes from both phages. Unlike DNA damage, induction via the newly discovered regulatory modules results in gene expression from and near-exclusive production of the phage responsive to the specific input. Our results suggest that the activities of these SOS-independent pathways dictate the outcomes of prophage-prophage competition by expanding the range of stimuli to which specific prophages can respond.

RESULTS

A bioinformatic search for SOS-inducible linear plasmid-like phages reveals phages that encode additional lysis-lysogeny regulatory modules.

76 To advance studies of inter-prophage competition in polylysogenic bacteria, we conducted a
77 search among sequenced phage genomes for genes encoding putative SOS-independent lysis-
78 lysogeny modules located between *repA* and *tefN*, hallmarks of all known linear plasmid-like
79 phages. Our strategy was inspired by the arrangement of the *vqmA_{Phage}* and *qtip* genes, which
80 are encoded between *repA* and *tefN* in vibriophage VP882.¹⁴ We searched all prophage genomes
81 on NCBI as well as six recently curated phage and phage-plasmid databases^{19–24} spanning
82 diverse environmental, marine, and human body sites for convergently-oriented *repA* and *tefN*
83 genes residing within 10 Kb of each other. The search revealed 784 putative linear plasmid-like
84 phage genomes, 274 of which contained unique yet conserved *repA-tefN* loci (Supplementary
85 Table 1). In 271 of 274 genomes (99%), the gene upstream of *repA* encodes a *cl*-like DNA-binding
86 protein adjacent to putative, divergently transcribed lytic, structural, and regulatory genes. A panel
87 of representative loci is shown in Figure 1a. We constrained our search to phage genomes
88 harboring RecA-dependent, autoproteolytic *cl* repressors because autoproteolytic repressors
89 exhibit a stereotypical response to SOS (i. e., repressor cleavage), which is bioinformatically
90 predictable and testable *in vitro*. Non-proteolytic *cl* repressors require additional phage- and/or
91 host-specific factors for regulation, thus, we did not consider them for the present study. For
92 phages possessing autoproteolytic *cl* repressors, we reasoned that any additional regulatory
93 modules uncovered in their genomes would likely respond to SOS-independent inputs. Filtering
94 phage genomes using these criteria led to 61 unique loci. The majority of phages eliminated at
95 this step (210/271) encoded *repA-tefN* loci resembling those of the *Escherichia coli* linear plasmid-
96 like phage called N15. This finding is likely a consequence of the overrepresentation of
97 *Enterobacteriaceae* in the sequencing databases. N15 is known to be subject to
98 antirepression,^{25,26} and we determined the repressor it encodes is non-proteolytic (Figure 1b). The
99 only member within the identified set of phages harboring autoproteolytic repressors that has
100 been previously investigated is phage VP882 (Figure 1a, b), described above, and a singleton in
101 our current cluster analysis (Supplementary Table 1). Many of the phages from our database

searches come from metagenomic sequencing projects, are unobtainable, and/or have no known host. Nonetheless, with the goal of probing the functions of these putative sensory input pathways, we focused on obtainable phage-host pairs.

Transcription factor-small ORF modules regulate prophage induction independently of SOS.

The first isolate we investigated is *Vibrio cyclitrophicus* 1F-97 (here forward called *Vibrio* 1F-97), which encodes a putative phage on genome contig 72, hereafter called phage 72. Between the phage 72 *repA* and *teIN* genes are genes encoding a putative transcription factor (as judged by its predicted DNA-binding domain according to the NCBI conserved domain database) and a counter-oriented small, 171 nt ORF (hereafter called TF₇₂ and smORF₇₂, respectively). The sequences of these genes lack nucleotide- and amino acid-level identity to the *vqmA_{Phage}* and *qtip* genes, however, their arrangement parallels that for *vqmA_{Phage}* and *qtip* in phage VP882. To test whether the DNA located between *repA* and *teIN* on phage 72 controls the phage lysis-lysogeny transition, smORF₇₂ was cloned under the control of an anhydrotetracycline (aTc)-inducible promoter on a plasmid (pTet-smORF₇₂) and conjugated into *Vibrio* 1F-97. Addition of aTc to this recombinant strain led to a precipitous decline in OD₆₀₀ similar to that when the potent SOS-activator ciprofloxacin was added, indicating phage-driven lysis occurred (Figure 1c). aTc administered to *Vibrio* 1F-97 harboring an empty vector, or to *Vibrio cholerae*, *Vibrio parahaemolyticus*, or *E. coli* (none of which contain phage 72) harboring the pTet-smORF₇₂ vector did not affect growth showing that smORF₇₂ protein alone is not lysis-inducing (Extended Data Figure 1a and 1b, respectively). We eliminated the possibility that an additional predicted 165 nt ORF encoded between *repA* and *teIN* functions similarly to smORF₇₂, as overexpression of the gene encoding this protein did not drive lysis (Figure 1a and Extended Data Figure 1c). Thus, host cell lysis requires the presence of the phage and induction of the smORF₇₂ gene.

Supplementation with ciprofloxacin did not activate *smORF*₇₂ expression in wild-type (WT) *Vibrio* 1F-97 (Extended Data Figure 1d), suggesting that *smORF*₇₂ production is SOS-independent. Importantly, phage preparations obtained from *Vibrio* 1F-97 treated with aTc or with ciprofloxacin contained phage 72 particles, indicating that phage 72 can be induced by both SOS-independent and SOS-dependent pathways (Extended Data Figure 1e).

We next explored how induction of *smORF*₇₂ promotes lysis. Important for this step is that our bioinformatic analysis revealed that 99% of the *repA-teIN* loci (271/274) contain genes upstream of *repA* that encode predicted phage repressor proteins (called *cl*₇₂ and its target promoter, *P*_{R72}, respectively, for phage 72). We verified the function of this repressor-promoter pair by fusing the *P*_{R72} promoter to *lux* on a plasmid (*P*_{R72}-*lux*). Recombinant *E. coli* carrying the construct made light, which decreased 500-fold when the gene encoding *cl*₇₂ was also introduced (Extended Data Figure 1f). Consistent with this result, purified *cl*₇₂ protein bound to the *P*_{R72} promoter *in vitro* (Extended Data Figure 1g). Introduction of pTet-*smORF*₇₂ into *E. coli* carrying the *cl*₇₂-*P*_{R72}-*lux* plasmid restored high level light production, indicating that production of *smORF*₇₂ inactivates the *cl*₇₂ repressor (Figure 1d). Unlike DNA damage, *smORF*₇₂ production did not lead to *cl*₇₂ proteolysis (Figure 1e). These results suggest that *smORF*₇₂ is an antirepressor that inactivates its partner *cl*₇₂ repressor protein via a nonproteolytic mechanism.

As noted above, a gene encoding a transcription factor, *tf*₇₂ lies adjacent to *smORF*₇₂. *TF*₇₂ is a logical candidate to be the regulator of *smORF*₇₂ expression. Indeed, production of *TF*₇₂ in *E. coli* harboring the *PsmORF*₇₂-*lux* plasmid increased light output by 28-fold compared to the empty vector (Figure 1f). Likewise, production of *TF*₇₂ in *Vibrio* 1F-97 led to an increase in *smORF*₇₂ transcription (Extended Data Figure 1h), after which the culture lysed (Figure 1g). These data suggest that *TF*₇₂, via transcriptional activation of *smORF*₇₂, drives phage 72-mediated lysis.

A central question is how *tf*₇₂ expression is naturally regulated in phage 72 to launch the smORF₇₂-mediated lysis cascade. In an effort to identify small molecule inducers of phage 72, we fused the promoter of *smORF*₇₂ to *lux* (*PsmORF*₇₂-*lux*) on a plasmid, introduced it into the *Vibrio* 1F-97 lysogen, and monitored both *PsmORF*₇₂-*lux* activity and growth (by OD₆₀₀) following exposure to various compounds. We tested commercially available compound libraries (Biolog MicroArrays, ~2,000 conditions) and a curated library of antibiotics (generous gift from the Seyedsayamdost Group, ~250 conditions). Among the antibiotic library compounds are known phage inducers including DNA-damaging agents and regulators of reactive oxygen species. A reporter for *E. coli* phage lambda induction (*P*_{Rlambda}-*lux*) was likewise assessed enabling us to determine if inducers of *PsmORF*₇₂-*lux* and/or inhibitors of growth were or were not phage 72-specific. No test compound elicited more than a two-fold increase in *PsmORF*₇₂-*lux* expression or a decline in OD₆₀₀ that was specific to phage 72 (Extended Data Figure 2a). Thus, our screens did not yield potential inducers. Indeed, identifying induction stimuli for prophages that are non-SOS-inducible has proven notoriously difficult.¹² Intriguingly, some phenothiazines, including common antipsychotic drugs, which have not previously been implicated in phage induction, drove low-level activation of both phage reporters (Extended Data Figure 2b).

Vibrio 1F-97 harbors at least two plasmid-like phages that control lysis via transcription factor-smORF modules encoded in different genomic regions

Bioinformatic searches (BLASTp) of the phage 72-encoded transcription factor-smORF module indicated that the closest homolog of the TF₇₂ regulatory protein in NCBI (~53% amino acid pairwise identity; Extended Data Figure 3a) resides on a different contig within the same *Vibrio* 1F-97 genome, contig 63. Contig 63, like phage 72, harbors signatures of a linear plasmid-like phage (e.g., genes analogous to *telN* and *repA* by synteny; Figure 2a). Furthermore, DNA

corresponding to contig 63 was present in phage preparations of ciprofloxacin-treated *Vibrio* 1F-97 cultures (Extended Data Figure 3b). We now refer to this element as phage 63, and we classify *Vibrio* 1F-97 as polylysogenic for phage 63 and phage 72. The two phages share little identity on a whole-genome basis (26.4% ANI), and all analogous genes, with the exception of the two transcription factors, share less than 30% amino acid identity (BLASTp). Furthermore, unlike phage 72, cloning and expression of the phage 63 DNA intervening *repA* and *tefN* did not induce lysis (Extended Data Figure 3c, d). Rather, the gene encoding the homologous transcription factor in phage 63 (*tf*₆₃) is located between genes encoding predicted partition machinery (*parAB*) and an operon encoding predicted late genes (tail and assembly genes; Figure 2a). *tf*₆₃ resides adjacent to, but in the opposite orientation from, a gene encoding another hypothetical 183 nt smORF (*smORF*₆₃). *smORF*₆₃ and *smORF*₇₂ have only 9 identical residues in pairwise alignment (11% amino acid identity (Extended Data Figure 3e)). We constructed the analogous set of tools described above for phage 72 for phage 62 and determined that the *par*-associated *tf*₆₃-*smORF*₆₃ pair encodes proteins that perform equivalent functions as the *tf*₇₂-*smORF*₇₂ pair (Extended Data Figure 4a-f). We were unable to identify a *tf*₆₃-*smORF*₆₃-specific small molecule inducer using the above compound screening strategy (Extended Data Figure 4g). We conclude that the *Vibrio* 1F-97 polylysogen harbors two plasmid-like phages, each with SOS-independent pathways to lysis.

To examine the specificities of the two TF-smORF lysis pathways, we assessed whether either phage TF could activate expression of the non-cognate *smORF* gene. Figure 2b shows that TF₇₂ did not cause light to be produced from a *PsmORF*₆₃-*lux* reporter, and TF₆₃ did not drive light production from the *PsmORF*₇₂-*lux* reporter. Interestingly, AlphaFold²⁷ structural predictions of TF₇₂ and TF₆₃ suggest the two TFs most closely resemble each other (RMSD = 1.3 Å) (Extended Data Figure 5a), restriction modification controller proteins (RMSD = 0.5 – 0.9 Å) and the helix-turn-helix (HTH) transcription factor, ClgR from *Mycobacterium smegmatis* (RMSD = 1.0 – 2.5 Å) (Extended Data Figure 5b). AlphaFold modeling pinpointed the residues in each phage TF that

comprise the DNA-recognition helix, $\alpha 3$ (Extended Figure 5a). Exchange of 5 residues within the recognition helix and flanking C-terminal loop reverses the TF₇₂ and TF₆₃ preferences for *smORF*₇₂ and *smORF*₆₃ promoter DNA (the chimeric proteins are called TF_{72-5x} and TF_{63-5x} Figure 2b and Extended Figure 3a). Thus, promoter specificity is conferred by a minimal set of non-contiguous unique residues in each TF. Consistent with the strict specificities of the phage pathway components, neither *smORF*, when co-produced with the non-cognate *cl* repressor (pTet-*smORF*₆₃ with *cl*₇₂-P_{R72}-*lux*; pTet-*smORF*₇₂ with *cl*₆₃-P_{R63}-*lux*), generated P_R-driven light production (Figure 2c). Thus, each phage-encoded TF-*smORF* module is specific at two levels: *smORF* expression and *smORF*-driven repressor inactivation.

A search for phages with genes adjacent to partition genes reveals homoserine lactone-quorum-sensing TF-smORF modules.

Our finding that phage genes encoding TF-*smORF* modules are not restricted to *repA-teIN* genomic regions motivated us to expand our search. Preliminary BLAST searches revealed putative TF₆₃ homologs in 117 predicted phage genomes, which were almost exclusively associated with the *Vibrionaceae* (Supplementary Table 2). In this dataset, we could identify TF₆₃ homologs in four major genomic contexts: (i) in *repA-teIN* loci, (ii) within 5 Kb of the partitioning gene *parB*, (iii) in a heterogeneous genomic region containing small, variable ORFs of unknown function and, (iv), most rarely, within 10 Kb of genes encoding predicted terminase and portal proteins. We used these findings to inform a second, “guilt-by-association”, search for putative lysis-controlling modules lacking detectable sequence homology to TF₆₃ beyond those encoded by *Vibrionaceae* revealed in the BLAST search.

NCBI was searched for sequences encoding homologs of ParB and the major capsid protein from phage 63. We chose to examine the *parB*-associated loci more deeply because this genomic context was one of the most common in our preliminary TF₆₃ BLAST search and contained a conserved indicator gene (*parB*), which other common context groups lacked (Supplementary Table 2). The ParB search was anchored against a phage capsid protein to enrich for phage-associated *parB* loci instead of homologous sequences in plasmids or bacterial genomes. Capsid genes represent some of the most conserved viral sequences and were common among the genomes containing *parB* in our original set of 117 phages with TF₆₃ homologs.²⁸ We filtered the output of this search for hits encoding an apparent TF within 10 Kb of the predicted *parB* gene, given the close association between TF₆₃ and ParB in our earlier query. This search strategy revealed 56 unique contigs (Supplementary Table 3), approximately 90% of which (50/56) are putative linear plasmid-like phages based on detectable *repA-teiN* loci. All 56 contigs were predicted to originate from phage genomes when analyzed by VIBRANT, a hybrid machine learning and protein similarity-based tool for phage identification (Supplementary Table 1).²⁹

In one *par*-associated node of interest containing 5 phages, each phage carries an operon containing genes encoding an XRE-like DNA-binding protein and a LuxR-type transcription factor (Figure 3a and Extended Data Figure 6a). LuxR-type proteins contain N-terminal acyl homoserine lactone (HSL) AI-binding domains and C-terminal helix-turn-helix DNA-binding domains.³⁰ The *luxR* genes for two of the identified phages (ARM81ld of *Aeromonas* sp. ARM81 and Apop of *Aeromonas popoffii*) are associated with host strains that we had previously identified in a bioinformatic search for phage-encoded LuxR proteins.³¹ In that work, we showed that the phage LuxR proteins could bind the HSLs *Aeromonads* are known to produce, however, we were unable to deduce the functions of the phage LuxR proteins at the time due to an inability to obtain an *Aeromonas* sp. ARM81 isolate with the phage and genetic intractability of *A. popoffii*. To expand

our understanding of the roles of these phage modules, we successfully obtained *Aeromonas* sp. ARM81 and we developed genetic tools to study *A. popoffii*.

First, we investigated the *A. popoffii* XRE-LuxR (XRE_{Apop}-LuxR_{Apop}) pair. An unannotated 156 nt *smORF* resides approximately 150 nucleotides upstream of the *xre_{Apop}-luxR_{Apop}* operon near the *par* locus (Extended Data Figure 6a). To determine if it is regulated by XRE_{Apop}-LuxR_{Apop}, a plasmid carrying a Tc-inducible *xre_{Apop}-luxR_{Apop}* was transformed into *E. coli* harboring *lux* fused to the promoter of the candidate *smORF* (*PsmORF_{Apop}-lux*). Light production from *PsmORF_{Apop}-lux* commenced only when the *E. coli* reporter strain was supplied with C4-HSL (Extended Data Figure 6b), an AI natively produced by *Aeromonads*.³² Thus, XRE_{Apop} and the LuxR_{Apop}-AI complex control *smORF* expression. Consistent with this finding, deletion of the *smORF* locus from Apop abolished XRE_{Apop}-LuxR_{Apop}-AI-dependent lysis of *A. popoffii* but not ciprofloxacin-dependent lysis (Extended Data Figure 6c). Analogous to what we show in Figure 1d and Extended Data Figures 1g, 4b, and 4f, the Apop cl protein, cl_{Apop}, shifts P_{R-Apop} DNA in vitro (Extended Figure 6d), and smORF_{Apop} inactivates cl_{Apop} (Extended Data Figure 6e). Lastly, unlike DNA damage, smORF_{Apop} did not lead to cl_{Apop} autoproteolysis (Extended Data Figure 6f). The *Aeromonas* sp. ARM81 phage ARM81ld XRE-LuxR module functions analogously to the *A. popoffii* XRE-LuxR module (Extended Data Figure 7a-e). Thus, these XRE-LuxR-controlled modules likely act as orthogonal SOS-independent pathways to induction.

The Apop and ARM81ld phage TF-smORF modules are unique among the characterized set of TF-smORF modules given that two TFs (XRE and LuxR) are apparently involved in activation of *smORF* expression. To define the individual and combined contributions of XRE_{Apop} and LuxR_{Apop}, we generated expression vectors carrying only pTet-*xre_{Apop}* or only pTet-*luxR_{Apop}* and tested them in the *PsmORF_{Apop}-lux* assay. Neither XRE_{Apop} nor LuxR_{Apop} alone drove reporter output irrespective of the presence of C4-HSL, indicating that in addition to AI, *smORF_{Apop}* expression

depends on both transcription factors (Figure 3b). To validate these findings, we engineered *xre*_{Apop}-*PsmORF*_{Apop}-*lux* and *luxR*_{Apop}-*PsmORF*_{Apop}-*lux* reporter plasmids, which carry the identical *PsmORF*_{Apop}-*lux* reporter sequence but that include either full-length *xre*_{Apop} or full-length *luxR*_{Apop}, controlled by the native promoter. We introduced pTet-*xre*_{Apop} or pTet-*luxR*_{Apop} into *E. coli* harboring the reporters. Heterologous *xre*_{Apop} expression induced light production only when *luxR*_{Apop} was present on the reporter plasmid and AI was supplied (Figure 3c, d). By contrast, heterologous *luxR*_{Apop} expression did not induce light production in either case, irrespective of the presence of AI (Figure 3c, d). We interpret these results as follows: XRE_{Apop} activates expression of the *xre*_{Apop}-*luxR*_{Apop} operon, driving both XRE_{Apop} and LuxR_{Apop} production. Together, LuxR_{Apop} bound to C4-HSL and XRE_{Apop}, are required to activate *smORF*_{Apop} expression, launching the phage lytic cascade and host-cell lysis (Figure 3e). Confirming the proposed regulatory arrangement *in vivo*, RT-qPCR revealed that overexpression of *xre*_{Apop} in *A. popoffii* led to rapid activation of expression of *xre*_{Apop} and *luxR*_{Apop}, whereas overexpression of *luxR*_{Apop} did not (Extended Data Figure 7f).

Phage-encoded transcription factor-small protein modules dictate which prophage is induced in a polylysogenic host.

Inter-prophage competition is predicted to arise in polylysogens when co-occurring prophages are induced by the same trigger (e.g., DNA damage).¹⁸ Our above discovery of non-canonical phage lysis pathways establishes the possibility that the outcomes of such competitions could depend on the particular inducer that launches lysis. *Vibrio* 1F-97 cells lyse following phage induction by ciprofloxacin (DNA damage), production of smORF₆₃, or production of smORF₇₂ (Figure 1c and Extended Data Figure 4e). Nonetheless, the amounts of phage 63 and phage 72 virions released under each condition could reflect the consequences of inter-prophage

competition. Whole genome sequencing of phage particle preparations from *Vibrio* 1F-97 cultures induced by ciprofloxacin, smORF₆₃, or smORF₇₂ revealed that ciprofloxacin treatment drove production of comparable quantities of DNA corresponding to both phage 72 and phage 63 (Figure 4a). In contrast, cultures induced by smORF₇₂ or smORF₆₃ resulted in near-exclusive production of phage 72 particles and phage 63 particles, respectively (Figure 4a). Thus, under non-SOS-inducing conditions, which phage is produced depends on the smORF driving lysis.

To date, investigations of prophage induction in polylysogenic hosts have been limited to population-level measurements, as in Figure 4a. Consequently, it is unclear whether, in a single host bacterium, co-residing prophages can be simultaneously induced and acquire the resources needed for replication, or alternatively, whether only one of the co-residing prophages successfully replicates. Specifically, in the case of the *Vibrio* 1F-97 polylysogen, multiple mechanisms could yield the mixture of phage 72 and phage 63 particles that are produced following DNA damage: First, each host cell could produce both phage 72 and phage 63 particles. Second, each host cell could produce exclusively phage 72 particles or exclusively phage 63 particles. Third, a subset of host cells could produce exclusively phage 72 particles, another subset could produce exclusively phage 63 particles, and a final subset could produce both phage 72 and phage 63 particles. To determine which possibility is correct, we used single molecule RNA-fluorescent in situ hybridization (smRNA-FISH) to mark and visualize phage induction in individual cells of the *Vibrio* 1F-97 polylysogen. Our strategy relied on the use of probes specific for phage 72 or for phage 63 genes that are expressed only during entry into the lytic cycle (see Methods). Figure 4b shows that following DNA damage, a subset of cells (33%) exclusively harbors RNA from phage 72, a subset (9%) harbors RNA from phage 63, and a subset (28%) harbors RNA from both phages. Thus, DNA damage causes heterogeneity in phage induction at the single cell level. By contrast, and as predicted based on the results in Figure 4a, only phage 72-specific RNA could be detected in cells induced by smORF₇₂, whereas only phage 63-specific

RNA could be detected in cells induced by smORF₆₃ (Figure 4b). These findings indicate that activation of TF-smORF modules drive exclusivity in prophage induction at the single-cell level. Conversely, SOS activation leads to multiple populations of differentially induced cells, including those expressing the lytic genes of both phages and of either phage alone.

Previous work determined that *Aeromonas* sp. ARM81 is polylysogenic and contains an integrated prophage, ARM81mr, in addition to the plasmid-like ARM81ld prophage.³³ We have no genomic evidence for an additional sensory pathway on the ARM81mr phage. We used the identical single-cell imaging smRNA-FISH analysis to discover the mechanism that gives rise to the sets of released phages from *Aeromonas* sp. ARM81 following induction with either ciprofloxacin or smORF_{ARM81ld}. Our results mirrored those for *Vibrio* IF-97. DNA damage resulted in cells expressing lytic genes of exclusively phage ARM81mr (36%), exclusively phage ARM81ld (13%), or both phages (46%), while smORF_{ARM81ld} production drove expression of only phage ARM81ld (Figure 4c). Indeed, quantitation of the virions released following treatment with ciprofloxacin or following induction of *xre*_{ARM81ld}-*luxR*_{ARM81ld} expression yielded results consistent with the smRNA-FISH analysis. Specifically, administration of ciprofloxacin led to a 140-fold increase in ARM81mr phage particles and a 22-fold increase in ARM81ld particles relative to when no inducer was added (Figure 4d). In contrast, induction of expression of *xre*_{ARM81ld}-*luxR*_{ARM81ld} with C4-HSL led to a 5-fold increase in ARM81ld particles, whereas ARM81mr particles were reduced 5-fold compared to their uninduced levels (amounting to a 700-fold reduction in ARM81mr particles compared to ciprofloxacin treatment; Figure 4d). Together, our results from *Vibrio* IF-97 and *Aeromonas* sp. ARM81 demonstrate that DNA damage results in induction of and competition between co-residing prophages, while possession of an additional sensory pathway drives host-cell lysis exclusively by the phage that encodes the module (Figure 4e). Thus, the particular induction cue dictates the distribution of phage particles produced by the

polylysogenic host. Moreover, single and dual phage production is possible from an individual host cell.

DISCUSSION

Despite most bacteria being predicted to be polylysogenic, an understanding of the consequences of polylysogeny on the hosts and on their resident phages has been constrained by the limited models available and a lack of known prophage induction cues beyond the SOS response. At present, competition among co-residing temperate prophages is generally considered a ‘sprint’ in which, in response to a single trigger (i.e., the SOS response), differences in particular phage properties (e.g., replication rates, packaging rates, burst size, etc.), dictate how many particles of each phage are produced. Our present work in polylysogens harboring phages with multiple pathways to lysis shows that differential phage induction can occur, and which phage or phages are produced varies depending on the induction cue. While tuning into the host SOS response remains a universal mechanism by which prophages perceive host-cell stress, the additional sensory pathways we discover here suggest that more specialized conditions exist which favor induction of one phage over another.

In this work, we focused on discovering and characterizing SOS-independent lysis induction pathways on linear plasmid-like phages that encode autoproteolytic *cl* repressors. We uncovered a pattern in which novel phage regulatory components are encoded adjacent to genes characteristic of linear plasmid-like phages (i.e., near replication and partition machinery genes). To make headway, we constrained our initial work to this subset of phages, which constitute a specific sub-type present in Gram-negative bacteria. We predict that SOS-independent pathways to lysis are widespread among phages but may require different search strategies to reveal them. Indeed, RecA-independent phage induction has been shown to occur in lambdoid phages in *E. coli*, however the molecular mechanisms underlying these alternative pathways are unknown.³⁴

380

381 Our single-cell imaging analyses further reveal that phage induction via the SOS cue drives a
382 heterogenous outcome from polylysogens. That is, individual cells produce one type of phage,
383 the other type of phage, or a mixture of both phages. In the cases in which only one phage's lytic
384 genes are expressed in a host cell, we speculate that perhaps some stochastic process causes
385 only one phage to enter the lytic cycle, or, alternatively, one phage achieves a sufficient "head
386 start" over the other phage and it monopolizes host resources, quenching production of the other
387 phage. Our finding that a mixture of phages can be produced by a single host cell shows that
388 multiple prophages can enter the lytic cycle and presumably compete for the host resources
389 required to replicate and produce viral particles. Regarding *Aeromonas*, we note that while the
390 smFISH analyses show that the SOS cue drives similar levels of early lytic gene induction from
391 both phages, phage ARM81mr ultimately outcompetes phage ARM81ld, producing 10-fold more
392 viral particles. Differences in rates of downstream processes (e.g., replication and/or packaging)
393 could drive this discrepancy. Our use of smRNA-FISH for imaging prophage induction in
394 polylysogenic bacteria enabled this first visualization of inter-prophage competition at the single-
395 cell level. Going forward, this approach can be easily adapted to study other polylysogens and,
396 for example, used for investigations of phage gene expression on rapid timescales during
397 prophage induction or infection.

398

399 Regarding the molecular mechanisms underlying inter-prophage competition, our results show
400 that induction of one phage's TF-smORF module does not cross-activate the lytic program of
401 another co-resident prophage. In *Vibrio* 1F-97, this specificity is achieved at two levels: specific
402 TF-mediated activation of *smORF* expression and specific smORF-driven repressor inactivation.
403 AlphaFold modeling of the TFs reveals similarity in the global folds of the two proteins; however,
404 remarkably, the exquisite specificity for their respective target promoters is conferred by a few key

residues within the DNA-recognition helix. In stark contrast to our findings, investigation of polylysogenic *Salmonella* strains revealed that some phage antirepressor proteins can inactivate cognate and non-cognate repressors, thus enabling synchronization of induction of multiple prophages.³⁵ This arrangement is thought to be vital for prophages with slow induction responses, allowing them to ‘piggyback’ off of prophages that are induced more rapidly.³⁵ The finding that prophage induction by an additional cue does not trigger induction of co-residing prophages could be a strategy that proves especially successful when host resources are limiting because it ensures exclusive reproduction and dissemination of only the induced phage.

The two phage-encoded LuxR proteins investigated here require partner XRE proteins to activate expression of their counter-oriented *smORF* genes, a requirement that does not exist for most bacterial LuxR QS receptors. One recent example, however, demonstrates that the activity of the *Pseudomonas aeruginosa* LuxR receptor, called RhlR, is modulated by direct interaction with the PqsE protein.³⁶ Whether the phage LuxR and XRE proteins interact directly, and if so, if the mechanism parallels that for RhlR-PqsE, remains to be determined. Finally, we found that *xre* and *luxR* do not always co-occur in phage genomes. A tBLASTn search using XRE_{Apop} as the query revealed three contigs, likely from linear plasmid-like phages, in different *Shigella sonnei* genomes (see Extended Data Figure 6a). Unlike in the *Aeromonas* phages, these putative *Shigella* phages lack *luxR* genes. *Shigella* are not known to produce HSL AIs, thus a LuxR component might not provide benefits to the phage in the *Shigella* host. Possibly, particular sets of regulatory components encoded on different plasmid-like phages have been shaped through evolution, presumably by relevant host-sensory cues.

The genes encoding the phage transcription factors TF₇₂, TF₆₃, XRE_{Apop}-LuxR_{Apop}, and XRE_{ARM81Id}-LuxR_{ARM81Id} all required synthetic induction to drive lysis, a situation that exactly mirrors our findings for VqmA_{Phage} in phage VP882.¹⁴ While not investigated here, it is possible that these

431 regulators are produced in infected cells and they function to activate lysis before any additional
432 newly infecting phages can establish lysogeny. Determining the roles of these modules in all
433 stages of the phage life cycle is a focus of our ongoing work. To date, we have not succeeded in
434 identifying the natural inducers of these TF-smORF cascades. However, we note that DNA-
435 damaging compounds, the universally employed phage inducers, do not generally occur in nature
436 at the concentrations used in experiments. Thus, the identities of the natural cues that drive lysis
437 even in intensively studied (i.e., SOS-dependent) prophages remain mysterious. Presumably, the
438 signals that induce the additional pathways to lysis revealed here, as well as others that are
439 identified going forward, could be particular to each host-phage partnership and the niche in which
440 they reside. We propose that, while discovered long after the SOS cue, some or all of these TF-
441 smORF pathways may in fact be key determinants of phage lifestyle transitions and the arbiters
442 of inter-prophage competition in real-world settings.

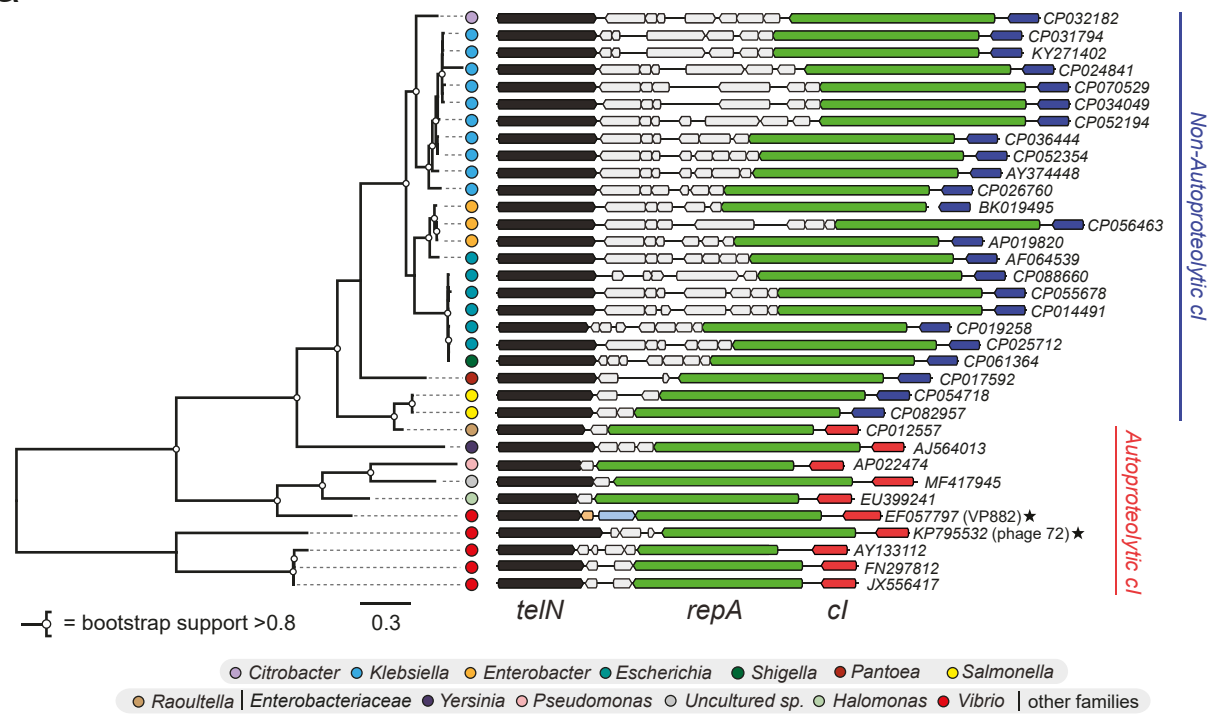
REFERENCES

- 444 1. Pires, D. P., Melo, L. D. R. & Azeredo, J. Understanding the Complex Phage-Host Interactions
445 in Biofilm Communities. *Annu. Rev. Virol.* **8**, 73–94 (2021).
- 446 2. Howard-Varona, C., Hargreaves, K. R., Abedon, S. T. & Sullivan, M. B. Lysogeny in nature:
447 mechanisms, impact and ecology of temperate phages. *ISME J.* **11**, 1511–1520 (2017).
- 448 3. Ofir, G. & Sorek, R. Contemporary Phage Biology: From Classic Models to New Insights. *Cell*
449 **172**, 1260–1270 (2018).
- 450 4. Ptashne, M. *A Genetic Switch, Third Edition: Phage Lambda Revisited*. (Cold Spring Harbor
451 Laboratory Press, 2004).
- 452 5. Lwoff, A. Lysogeny. *Bacteriol. Rev.* **17**, 269–337 (1953).
- 453 6. Roux, S. *et al.* Cryptic inoviruses revealed as pervasive in bacteria and archaea across
454 Earth's biomes. *Nat. Microbiol.* **4**, 1895–1906 (2019).
- 455 7. Hay, I. D. & Lithgow, T. Filamentous phages: masters of a microbial sharing economy. *EMBO*
456 *Rep.* **20**, e47427 (2019).
- 457 8. Sauer, R. T., Ross, M. J. & Ptashne, M. Cleavage of the lambda and P22 repressors by recA
458 protein. *J. Biol. Chem.* **257**, 4458–4462 (1982).
- 459 9. Little, J. W. Autodigestion of lexA and phage lambda repressors. *Proc. Natl. Acad. Sci.* **81**,
460 1375–1379 (1984).
- 461 10. Shearwin, K. E., Brumby, A. M. & Egan, J. B. The Tum Protein of Coliphage 186 Is an
462 Antirepressor. *J. Biol. Chem.* **273**, 5708–5715 (1998).
- 463 11. Bose, B., Auchtung, J. M., Lee, C. A. & Grossman, A. D. A conserved anti-repressor controls
464 horizontal gene transfer by proteolysis. *Mol. Microbiol.* **70**, 570–582 (2008).
- 465 12. Campbell, D. E. *et al.* Infection with Bacteroides Phage BV01 Alters the Host Transcriptome
466 and Bile Acid Metabolism in a Common Human Gut Microbe. *Cell Rep.* **32**, 108142 (2020).
- 467 13. Hargreaves, K. R., Kropinski, A. M. & Clokie, M. R. J. What Does the Talking?: Quorum
468 Sensing Signalling Genes Discovered in a Bacteriophage Genome. *PLoS ONE* **9**, (2014).
- 469 14. Silpe, J. E. & Bassler, B. L. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage
470 Lysis-Lysogeny Decision. *Cell* **176**, 268–280.e13 (2019).
- 471 15. Erez, Z. *et al.* Communication between viruses guides lysis–lysogeny decisions. *Nature* **541**,
472 488–493 (2017).
- 473 16. Papenfort, K. & Bassler, B. L. Quorum sensing signal–response systems in Gram-negative
474 bacteria. *Nat. Rev. Microbiol.* **14**, 576–588 (2016).
- 475 17. Papenfort, K. *et al.* A *Vibrio cholerae* autoinducer–receptor pair that controls biofilm formation.
476 *Nat. Chem. Biol.* **13**, 551–557 (2017).
- 477 18. Refardt, D. Within-host competition determines reproductive success of temperate
478 bacteriophages. *ISME J.* **5**, 1451–1460 (2011).
- 479 19. Tisza, M. J. & Buck, C. B. A catalog of tens of thousands of viruses from human metagenomes
480 reveals hidden associations with chronic diseases. *Proc. Natl. Acad. Sci.* **118**, e2023202118
481 (2021).
- 482 20. Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D.
483 Massive expansion of human gut bacteriophage diversity. *Cell* **184**, 1098–1109.e9 (2021).
- 484 21. Pfeifer, E., Moura de Sousa, J. A., Touchon, M. & Rocha, E. P. C. Bacteria have numerous
485 distinctive groups of phage–plasmids with conserved phage and variable plasmid gene
486 repertoires. *Nucleic Acids Res.* **49**, 2655–2673 (2021).
- 487 22. Roux, S. *et al.* Ecogenomics and potential biogeochemical impacts of globally abundant
488 ocean viruses. *Nature* **537**, 689–693 (2016).
- 489 23. Roux, S. *et al.* IMG/VR v3: an integrated ecological and evolutionary framework for
490 interrogating genomes of uncultivated viruses. *Nucleic Acids Res.* **49**, D764–D775 (2021).

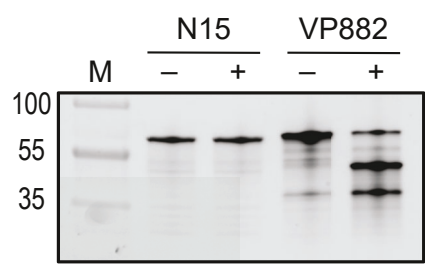
- 491 24. Nayfach, S. *et al.* Metagenomic compendium of 189,680 DNA viruses from the human gut
492 microbiome. *Nat. Microbiol.* **6**, 960–970 (2021).
- 493 25. Ravin, N. V., Svarchevsky, A. N. & Dehò, G. The anti-immunity system of phage-plasmid N15:
494 identification of the antirepressor gene and its control by a small processed RNA. *Mol.*
495 *Microbiol.* **34**, 980–994 (1999).
- 496 26. Mardanov, A. V. & Ravin, N. V. The Antirepressor Needed for Induction of Linear Plasmid-
497 Prophage N15 Belongs to the SOS Regulon. *J. Bacteriol.* **189**, 6333–6338 (2007).
- 498 27. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–
499 589 (2021).
- 500 28. Koonin, E. V. *et al.* Global Organization and Proposed Megataxonomy of the Virus World.
501 *Microbiol. Mol. Biol. Rev.* **84**, e00061-19 (2020).
- 502 29. Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and
503 curation of microbial viruses, and evaluation of viral community function from genomic
504 sequences. *Microbiome* **8**, 90 (2020).
- 505 30. Fuqua, W. C., Winans, S. C. & Greenberg, E. P. Quorum sensing in bacteria: the LuxR-LuxI
506 family of cell density-responsive transcriptional regulators. *J. Bacteriol.* **176**, 269–275 (1994).
- 507 31. Silpe, J. E. & Bassler, B. L. Phage-Encoded LuxR-Type Receptors Responsive to Host-
508 Produced Bacterial Quorum-Sensing Autoinducers. *mBio* **10**, (2019).
- 509 32. Silpe, J. E., Duddy, O. P. & Bassler, B. L. Natural and synthetic inhibitors of a phage-encoded
510 quorum-sensing receptor affect phage–host dynamics in mixed bacterial communities. *Proc.*
511 *Natl. Acad. Sci.* **119**, e2217813119 (2022).
- 512 33. Dziewit, L. & Radlinska, M. Two novel temperate bacteriophages co-existing in *Aeromonas*
513 sp. ARM81 – characterization of their genomes, proteomes and DNA methyltransferases. *J.*
514 *Gen. Virol.* **97**, 2008–2022 (2016).
- 515 34. Rozanov, D. V., D’Ari, R. & Sineoky, S. P. RecA-Independent Pathways of Lambdoid
516 Prophage Induction in *Escherichia coli*. *J. Bacteriol.* **180**, 6306–6315 (1998).
- 517 35. Lemire, S., Figueroa-Bossi, N. & Bossi, L. Bacteriophage Crosstalk: Coordination of
518 Prophage Induction by Trans-Acting Antirepressors. *PLoS Genet.* **7**, e1002149 (2011).
- 519 36. Simanek, K. A. *et al.* The PqsE-RhlR Interaction Regulates RhlR DNA Binding to Control
520 Virulence Factor Production in *Pseudomonas aeruginosa*. *Microbiol. Spectr.* **10**, e02108-21
521 (2022).
- 522
- 523

Figure 1

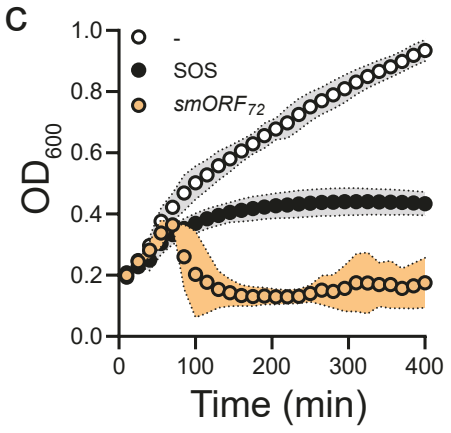
a



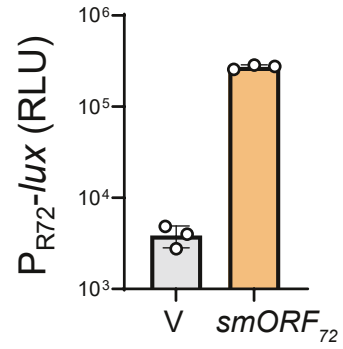
b



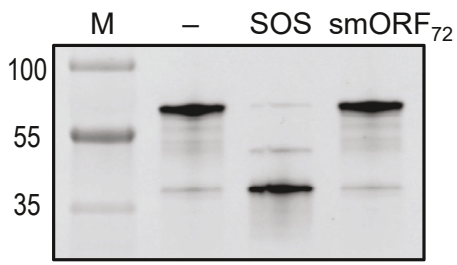
c



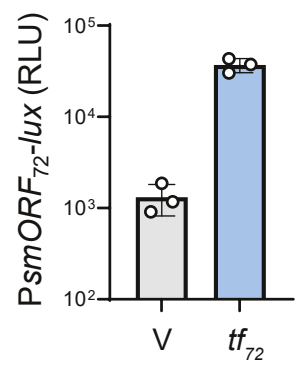
d



e



f



g

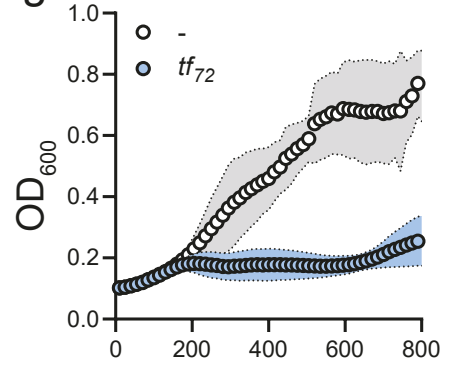


FIGURE LEGENDS

Figure 1. Variable gene content in an otherwise conserved locus of linear plasmid-like phages reveals TF-smORF modules that regulate lysis independently of SOS.

(a) Phylogenetic tree (left) of 34 representative *TeIN* proteins. *TeIN* protein groups typically segregate with bacterial genera (colored circles, legend at bottom). Nodes with bootstrap support >0.8 are indicated with white circles. Gene neighborhoods (right) for the 34 loci encoding convergently-oriented *teIN* (black) and *repA* (green) genes. The variable loci encoding *tf-smORF* modules are located between *teIN* and *repA*. Phage VP882 and *Vibrio* 1F-97 phage 72 are denoted with stars and the SOS-independent pathway components in phage VP882 (*vqmA_{Phage}* and *qtip*) are colored blue and orange, respectively. All other genes are colored gray, encode unknown functions, and vary across loci. Genes encoding predicted autoproteolytic repressors (red) cluster together with respect to *TeIN* phylogeny and are distinct from genes encoding non-cleavable repressors (navy). NCBI accession numbers are depicted to the right of each sequence. The scale bar indicates the number of amino acid substitutions per site.

(b) SDS-PAGE in-gel labeling of the non-proteolytic N15 phage repressor (HALO-cl_{N15}) and the autoproteolytic phage VP882 repressor (HALO-cl_{VP882}). - and + indicate, respectively, the absence and presence of 500 ng mL⁻¹ ciprofloxacin used to induce the SOS response. M denotes the molecular weight marker (representative bands are labeled).

(c) Growth of *Vibrio* 1F-97 carrying aTc-inducible *smORF₇₂* in medium containing 50 ng mL⁻¹ aTc (denoted *smORF₇₂*, orange), 500 ng mL⁻¹ ciprofloxacin (denoted SOS, black), or water (denoted -, white).

(d) *P_{R72-lux}* expression in *E. coli* carrying an empty vector (designated V) or aTc-inducible *smORF₇₂* grown in medium containing aTc. The *P_{R72-lux}* plasmid carries *cl₇₂*, which natively represses reporter expression. Relative light units (RLU) were calculated by dividing bioluminescence by OD₆₀₀. aTc concentration as in (c).

(e) SDS-PAGE in-gel labeling of the *Vibrio* 1F-97 phage 72 repressor (cl₇₂-HALO) produced by *E. coli* carrying aTc-inducible *smORF₇₂*. The treatments -, SOS, and *smORF₇₂* refer to water, ciprofloxacin, and aTc, respectively, with concentrations as in (c). M as in (b).

(f) *PsmORF_{72-lux}* expression from *E. coli* carrying an empty vector (V) or aTc-inducible *tf₇₂* grown in medium containing aTc. RLU as in (d). aTc concentration as in (c).

(g) Growth of *Vibrio* 1F-97 carrying aTc-inducible *tf₇₂* in medium lacking or containing aTc (white and blue, respectively). aTc concentration as in (c).

Data are represented as means ± std with *n*=3 biological replicates (c, d, f, g) and as a single representative image (b, e).

Figure 2

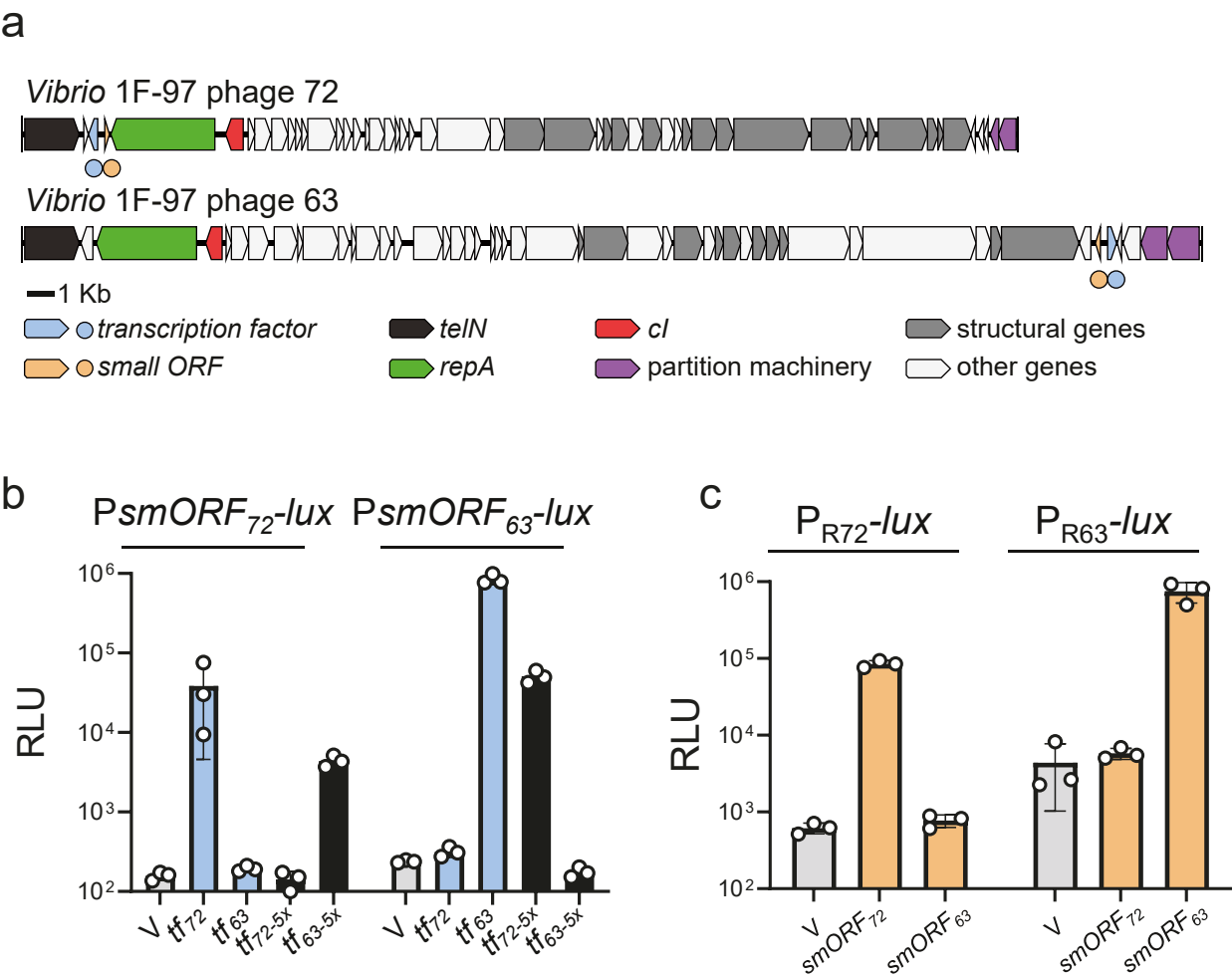


Figure 2. *Vibrio* 1F-97 harbors two linear plasmid-like phages that control lysis via TF-smORF modules encoded in different genomic loci.

(a) Organization of genes on linear plasmid-like phages in *Vibrio* 1F-97. Genes are colored by functional annotation as noted in the key. Circles below genomes denote positions of the *tf* (blue) and *smORF* (orange) genes located at *repA-telN* (phage 72) or by *par* genes (phage 63). All analogous genes, with the exception of those encoding the TFs, share less than 30% amino acid identity (BLASTp).

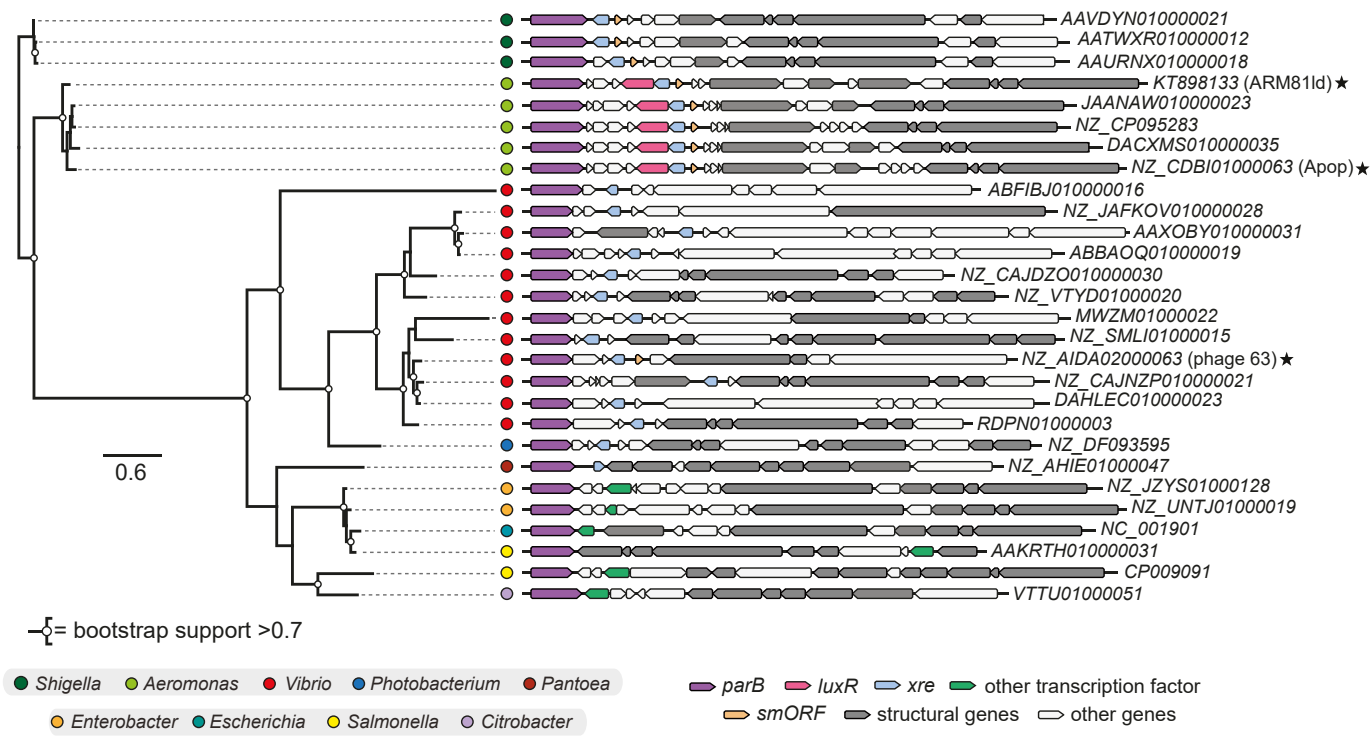
(b) *PsmORF-lux* output from *E. coli* harboring *PsmORF*₇₂-*lux* or *PsmORF*₆₃-*lux* and a second plasmid encoding either an empty vector (V), aTc-inducible *tf*₇₂, *tf*₆₃, or the chimeric *tf* genes, designated *tf*_{72-5x} and *tf*_{63-5x}, each grown in medium containing aTc. The chimeric TF proteins (black bars) have 5 amino acids in their DNA binding domains replaced from the other TF. Thus, TF_{72-5x} possesses 5 amino acids from TF₆₃ and TF_{63-5x} possesses 5 amino acids from TF₇₂. Extended Data Figure 3a shows the exchanged residues and their locations.

(c) Light production from *E. coli* harboring cl-repressed P_{R72}-*lux* or P_{R63}-*lux* reporter plasmids and a second plasmid encoding either an empty vector (V), aTc-inducible *smORF*₇₂ or aTc-inducible *smORF*₆₃, each grown in medium containing aTc.

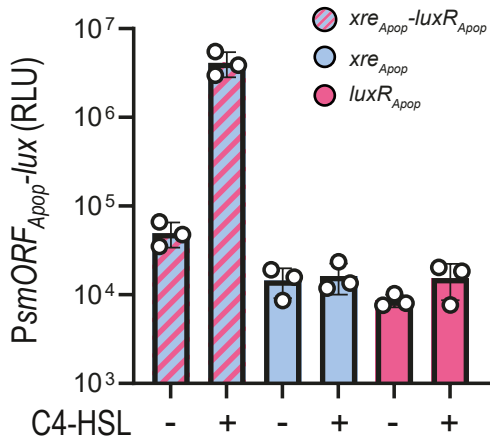
Data are represented as means ± std with *n*=3 biological replicates (b, c). RLU as in Figure 1d (b, c). aTc; 50 ng mL⁻¹ (b, c).

Figure 3

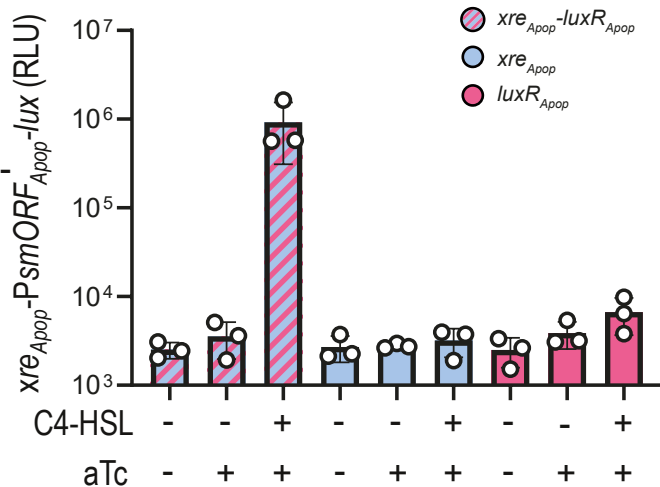
a



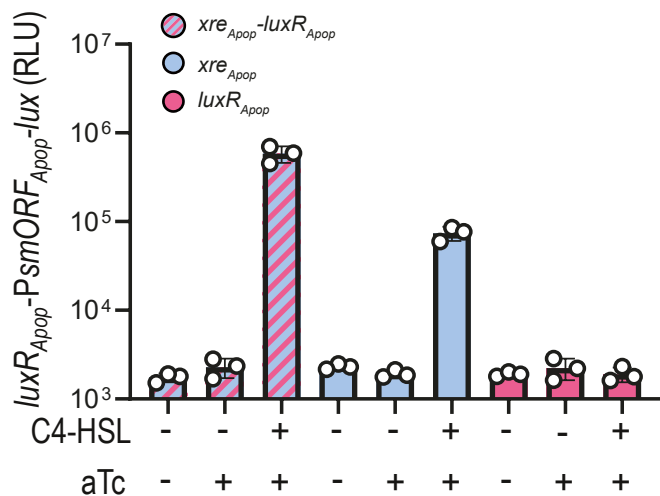
b



c



d



e

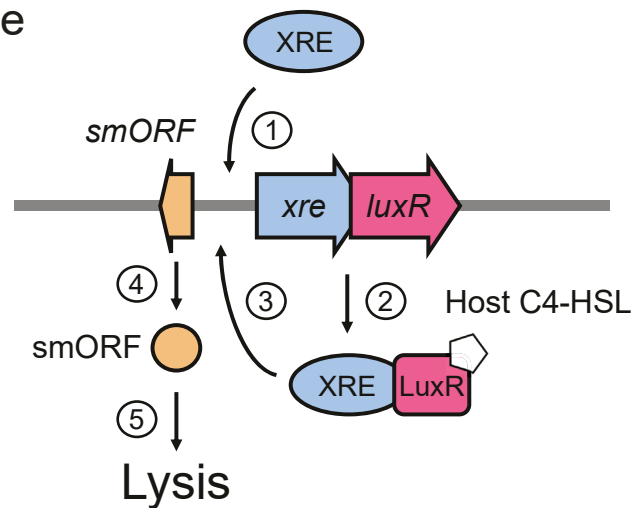


Figure 3. Activation of *PsmORF*_{Apop} requires *XRE*_{Apop}, *LuxR*_{Apop}, and the C4-HSL AI.

(a) Phylogenetic tree (left) of 28 representative ParB proteins. ParB protein groups typically segregate with bacterial genera (colored circles, legend at bottom), but these groups are not always concordant with deeper taxonomic relationships. Nodes with bootstrap support >0.7 are indicated with white circles. Gene neighborhoods (right) for the 28 loci encoding *parB* (purple) and the major capsid protein (gray). ARM81Id, Apop, and phage 63 are denoted with stars. The *luxR*, *xre*, and *smORF* genes are colored pink, blue, and orange, respectively. Genes encoding predicted transcription factors that are not *luxR* or *xre* are colored in green. NCBI accession numbers are depicted to the right of each sequence. The scale bar indicates the number of amino acid substitutions per site.

(b) *PsmORF*_{Apop}-*lux* activity from a plasmid in *E. coli* carrying a second vector with the designated aTc-inducible gene(s). All media contained aTc. Supplementation with DMSO or C4-HSL is denoted – and +, respectively.

(c) *PsmORF*_{Apop}-*lux* activity from a plasmid in *E. coli* carrying *xre*_{Apop} under its native promoter and a second vector with the designated aTc-inducible gene(s). Media contained the indicated combinations of water (-) or aTc (+), and DMSO (-) or C4-HSL (+).

(d) *PsmORF*_{Apop}-*lux* activity from a plasmid in *E. coli* carrying *luxR*_{Apop} under its native promoter and a second vector with the designated aTc inducible gene(s). Treatments as in (c).

(e) Proposed model for regulation of the TF-smORF module in Apop. (1) *XRE*_{Apop} activates expression of the *xre*_{Apop}-*luxR*_{Apop} operon, and (2) increased production of *LuxR*_{Apop} and *XRE*_{Apop} occurs. (3) *LuxR*_{Apop} when bound to the C4-HSL AI ligand (pentagon), together with *XRE*_{Apop}, activates expression of the counter-oriented *smORF*_{Apop} gene, and also activates expression of *xre*_{Apop}-*luxR*_{Apop}. (4) *smORF*_{Apop} inhibits the *cl*_{Apop} repressor, (5) leading to host-cell lysis. The co-requirement for AI-bound *LuxR*_{Apop} and *XRE*_{Apop} could be due to a protein-protein interaction or an alteration in the topology of the *XRE*_{Apop} DNA-binding site in favor of expression of *smORF*_{Apop}. The mechanism remains to be determined.

Data are represented as means ± std with *n*=3 biological replicates (b, c, d). RLU as in Figure 1d (b, c, d). aTc; 50 ng mL⁻¹ (b, c, d), C4-HSL; 10 μM (b, c, d).

Figure 4

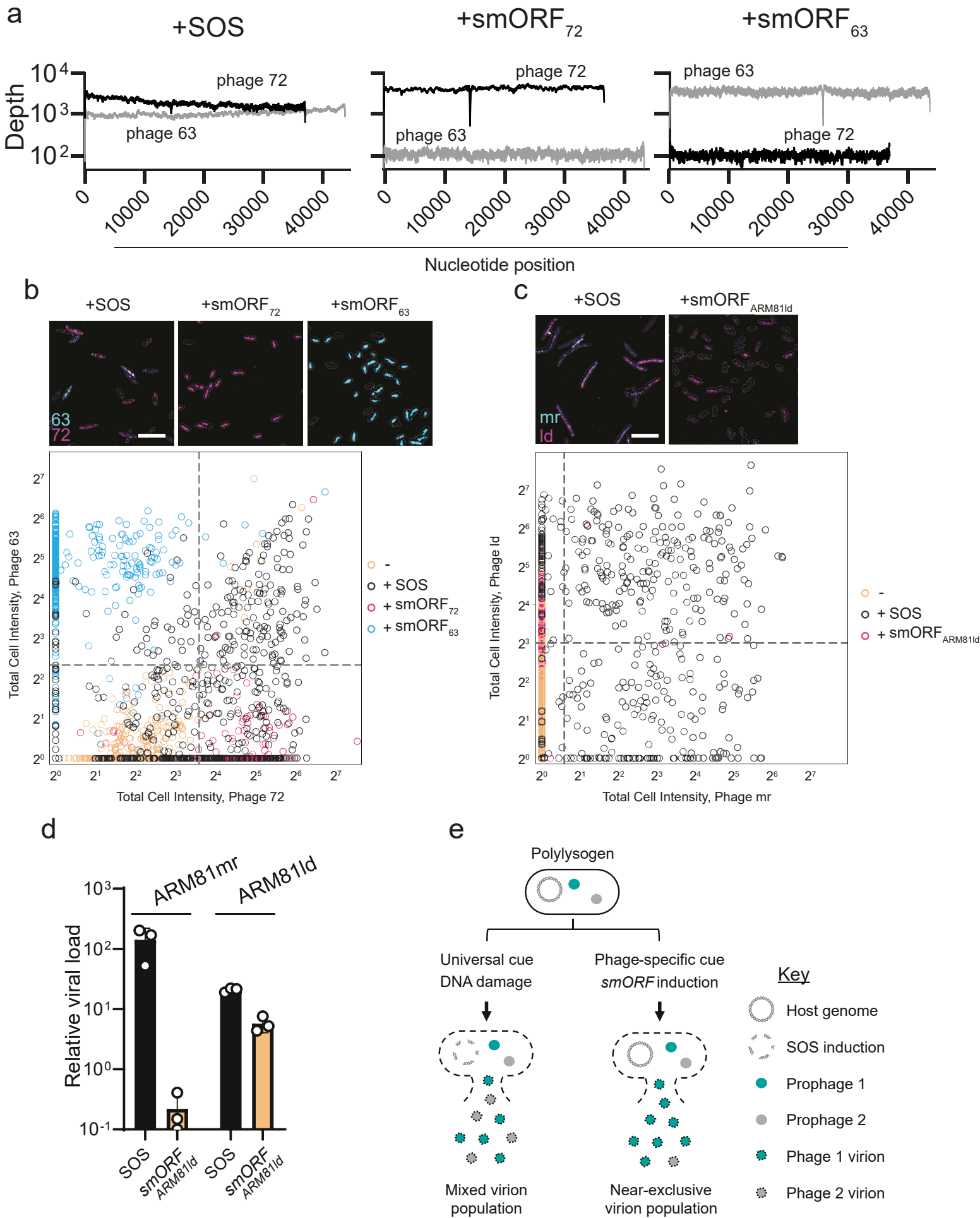


Figure 4. DNA-damage drives discrete subsets of cells producing one, the other, or both phages while TF-smORF modules exclusively drive production of only the phage encoding that module in polylysogenic hosts.

(a) Whole-genome sequencing of viral particles prepared from *Vibrio* 1F-97 cultures carrying an empty vector, a vector with aTc-inducible *smORF*₇₂, or a vector with aTc-inducible *smORF*₆₃ grown in medium containing ciprofloxacin for activation of the SOS response or aTc for *smORF* induction. Depth refers to the number of read counts. Phage 72 (black) is 43.1 Kb and phage 63 (gray) is 36.5 Kb in genome size.

(b) Top: Representative smFISH images of phage 63 and phage 72 early lytic gene expression in *Vibrio* 1F-97 cells following activation of the SOS response or *smORF* induction as described in (a). Images are maximum z-projections of raw smFISH fluorescence for phage 72 genes (magenta) and phage 63 genes (cyan) with cells outlined in white. Bottom: Total phage 72 versus phage 63 smFISH intensity per *Vibrio* 1F-97 cell in the absence of phage induction (orange, n=677 cells), with SOS activation (black, n=516 cells), *smORF*₇₂ induction (magenta, n=375 cells), or *smORF*₆₃ induction (cyan, n=485 cells). Uninduced cells were used to determine boundaries (gray dotted lines, see Methods) to delineate cells displaying no phage induction (bottom left quadrant), exclusively phage 72 induction (bottom right quadrant), exclusively phage 63 induction (top left quadrant), or both phage 63 and phage 72 induction (top right quadrant). Scale bar denotes 10 μ m.

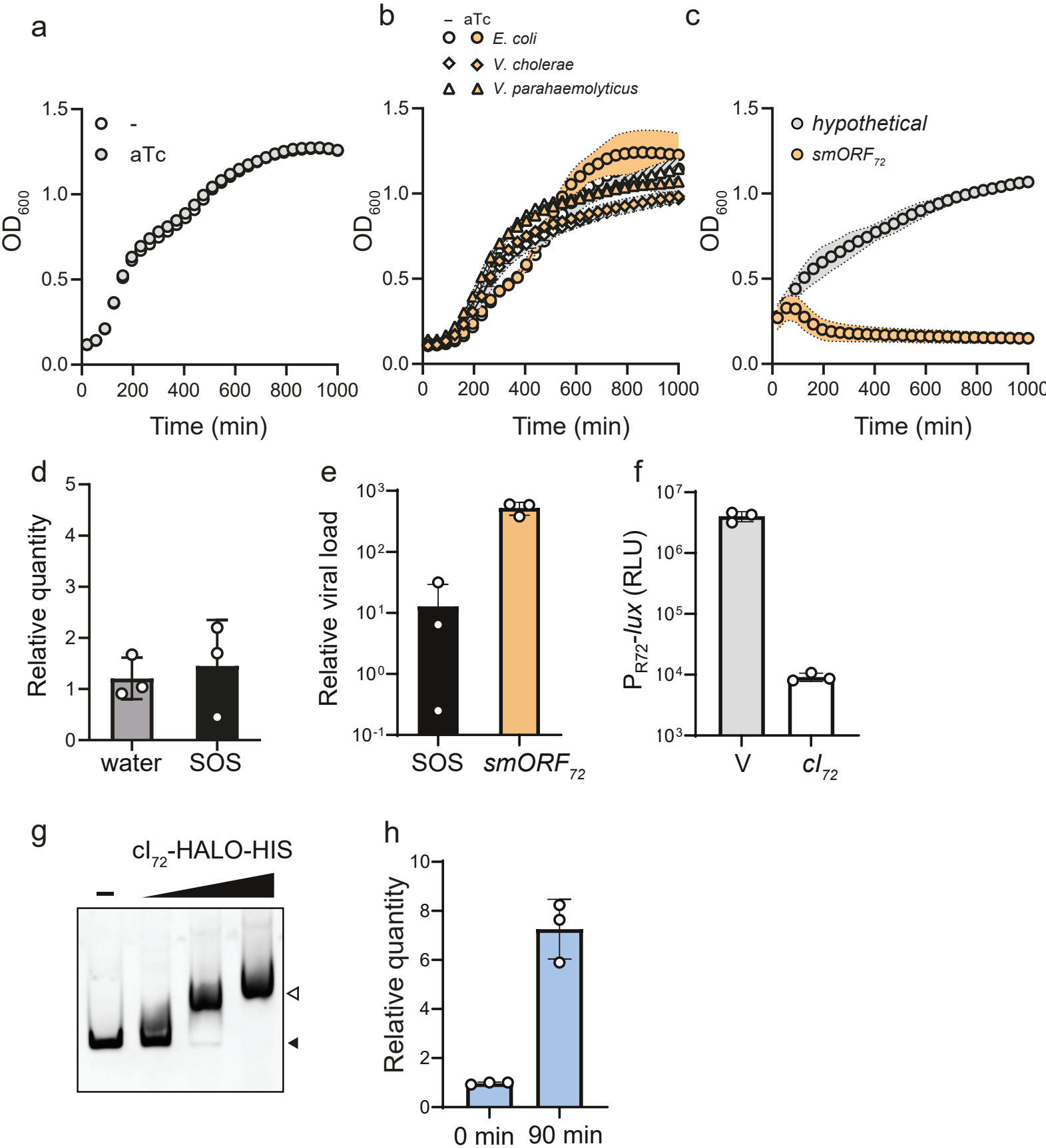
(c) Top: Representative smFISH images of phage ARM81mr and phage ARM81ld early lytic gene expression in *Aeromonas* sp. cells following activation of the SOS response or *smORF*_{ARM81ld} induction as described in (d). Bottom: Images as in (b) for phage ARM81mr genes (cyan) and phage ARM81ld genes (magenta). Total phage ARM81mr versus phage ARM81ld smFISH intensity per *Aeromonas* sp. cell in the absence of phage induction (orange, n=516 cells), with SOS activation (black, n=463 cells), or *smORF*_{ARM81ld} induction (magenta, n=477 cells). Orange and magenta points frequently overlap and are concentrated on the vertical axis. Scale bar and gray dotted lines delineating quadrants (see Methods) as in (b). Cells displaying no phage induction, exclusively phage ARM81mr induction, exclusively phage ARM81ld induction, or both phage ARM81mr and phage ARM81ld induction are represented in the bottom left, bottom right, top left, and top right quadrants, respectively.

(d) Detection of viral particles prepared from *Aeromonas* sp. ARM81 cultures carrying aTc-inducible *smORF*_{ARM81ld} grown in medium containing ciprofloxacin (black) or aTc (orange) to induce SOS or *smORF*_{ARM81ld} production, respectively. All media contained C4-HSL. Relative viral load is measured as the amount of ARM81mr- or ARM81ld-specific DNA (*cl*_{ARM81mr} and *cl*_{ARM81ld}, respectively) in the induced samples relative to an uninduced sample measured by qPCR.

(e) Proposed model for how TF-smORF modules influence inter-prophage competition outcomes in polylysogenic bacteria. Left side: Exposure of a polylysogen to a ubiquitous trigger (i.e., SOS activation/DNA damage) fosters competitive conditions in which multiple prophages share host cell resources to replicate and produce viral particles. Right side: Prophage-specific induction via a TF-smORF module (the inducer of the teal prophage is present in this example) leads to non-competitive conditions in which only the induced phage garners host cell resources leading to near exclusive production of particles of only that phage.

Data are represented as means $\pm$ std with *n*=3 biological replicates (a, b, c) and as means $\pm$ std with *n*=3 biological replicates and *n*=4 technical replicates (d). aTc; 50 ng mL⁻¹ (a), aTc; 5 ng mL⁻¹ (d), ciprofloxacin; 500 ng mL⁻¹ (a), ciprofloxacin; 1 μ g mL⁻¹ (d), C4-HSL; 10 μ M (d).

Extended Data Figure 1

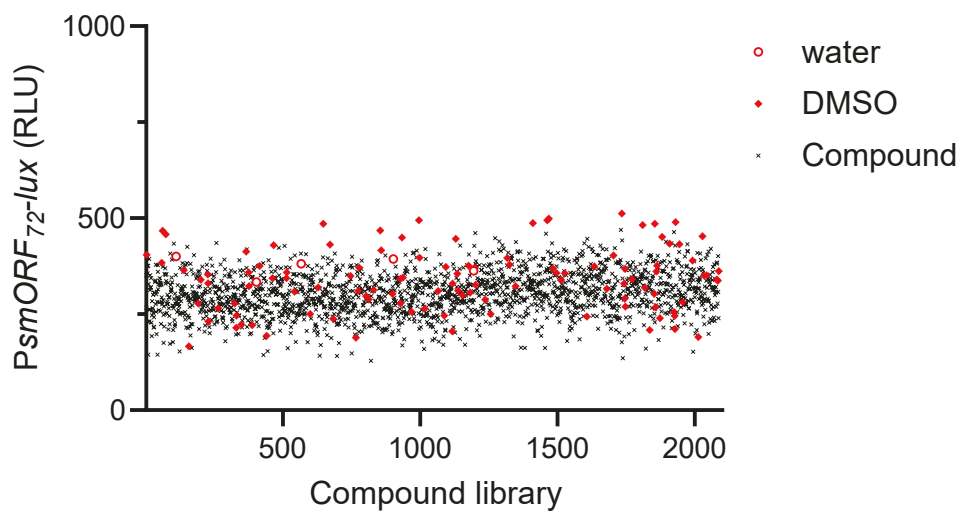


Extended Data Figure 1. smORF₇₂-induced lysis of *Vibrio* 1F-97 is phage dependent and requires a TF-smORF module.

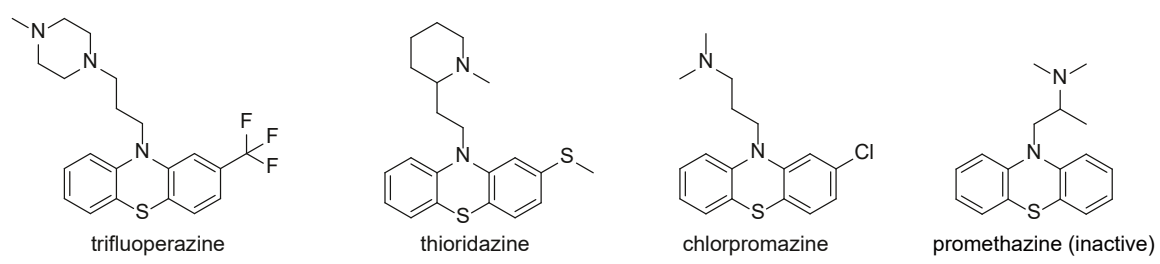
- (a) Growth of WT *Vibrio* 1F-97 in medium lacking or containing aTc (white and gray, respectively).
- (b) Growth of *E. coli* (circles), *V. cholerae* (diamonds), and *V. parahaemolyticus* (triangles) carrying aTc-inducible smORF₇₂ in medium lacking or containing aTc (white and orange, respectively).
- (c) Growth of WT *Vibrio* 1F-97 carrying an aTc-inducible smORF gene (designated *hypothetical*) that resides between *repA-teIN* on phage 72 or aTc-inducible smORF₇₂ in medium containing aTc (gray and orange, respectively).
- (d) Relative expression of smORF₇₂ as judged by RT-qPCR in *Vibrio* 1F-97 15 min after addition of water or ciprofloxacin. Relative transcript levels are the amount of phage smORF₇₂ RNA relative to the amount of *rpoB* RNA, normalized to T=0 min.
- (e) Detection of phage 72-specific particles in viral preparations obtained from culture fluids from *Vibrio* 1F-97 carrying aTc-inducible smORF₇₂ that were grown in medium with ciprofloxacin or aTc to induce SOS and smORF₇₂ production, respectively. Relative viral load is the amount of *cl*₇₂ DNA in the induced samples relative to that in an uninduced sample as judged by qPCR.
- (f) Expression of plasmid-borne P_{R72}-*lux* in *E. coli* containing an empty vector (V) or the phage *cl*₇₂ gene.
- (g) EMSA showing binding of *cl*₇₂-HALO-HIS protein to P_{R72} DNA. Approximately 14 nM of P_{R72} DNA was combined with 800, 1600, or 3200 nM of *cl*₇₂-HALO-HIS protein. The no protein control lane is designated with a minus sign. Locations of the unshifted and shifted probe are indicated with black and white arrows, respectively.
- (h) Relative expression of smORF₇₂ as judged by RT-qPCR in *Vibrio* 1F-97 carrying aTc-inducible TF₇₂ at 0 min and 90 min after induction with aTc. Relative transcript levels are the amount of phage smORF₇₂ RNA relative to the amount of *rpoB* RNA, normalized to T = 0 min. Data are represented as means ± std with *n*=3 biological replicates (a, b, c, e, f), as a single representative image (g), and as means ± std with *n*=3 biological replicates and *n*=4 technical replicates (d, h). RLU as in Figure 1d (f). aTc; 50 ng mL⁻¹ (a, b, c, e, h), ciprofloxacin; 500 ng mL⁻¹ (d, e).

Extended Data Figure 2

a



b



Extended Data Figure 2. No *tf₇₂-smORF₇₂*-specific inducer was identified in two small-molecule screens.

(a) Relative *PsmORF₇₂-lux* expression in *Vibrio* 1F-97 cultured in Biolog microarray plates and with a curated library of antibiotics. The red circles and diamonds show the water and DMSO vehicle controls, respectively. The black x symbols show the results for the different compounds tested. Figure 1f shows assessment of reporter function. Conditions that did not support growth were excluded from analysis. Data are represented as a single reading. RLU as in Figure 1d.

(b) Structures of phenothiazines identified from the compound library that are not known DNA-damaging agents and that induce a phage lambda-derived *cl-P_R-lux* reporter (see Methods).

Extended Data Figure 3

a

Consensus M X K X D P I V X V L R X X R E X X G X X Q X X X G X T G J S X X T I Q R X E X G X T D M K L X X Y R X X L B A L X X X D M D V S X A L X S H E F X X X X

TF₇₂ 1 M K Y D P I V S V L R H C R E A K G M T Q K D L A G Y T G L S Y R T I Q R M E S G D T D M K L G Y R R I L D A L D L T D M D V S V A L L S H E F T K A D

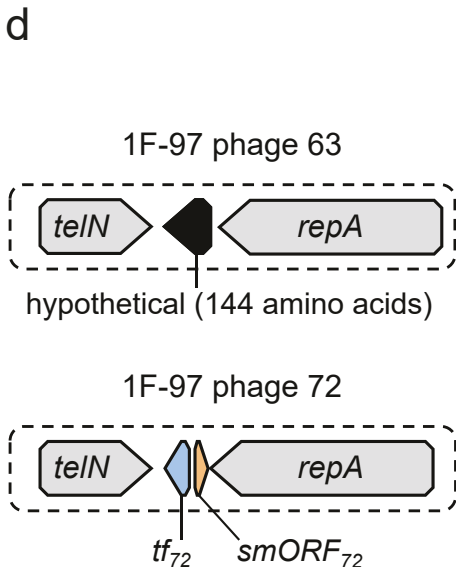
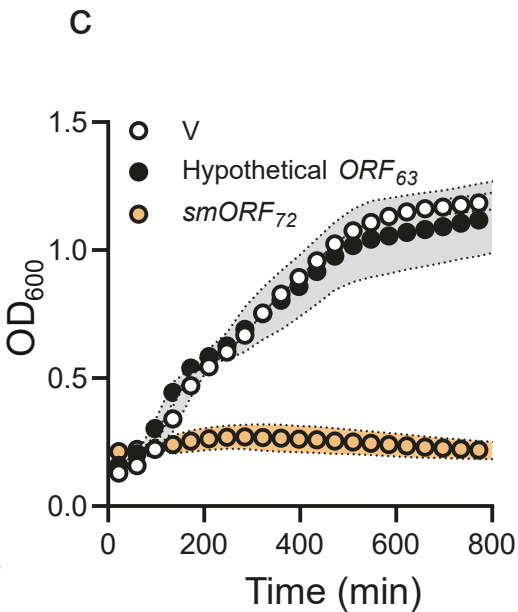
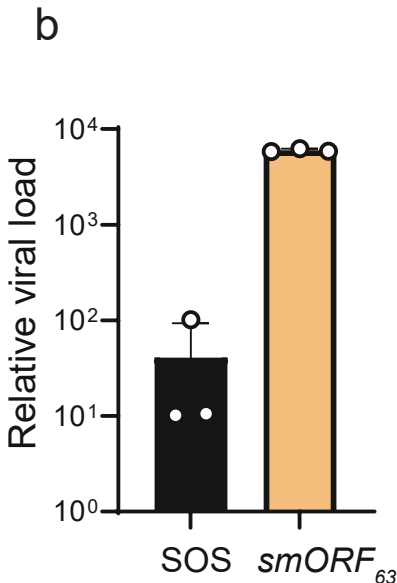
TF₆₃ 1 M K K A D P I V Y V L R R Y R E N Q G I S Q D K M S G L T G I S V S T I Q R I E N G R T D M K L S H Y R S Y L N A L G M S D M D V S I A L F S H E F V T E K

★★ ★ ★

Consensus D V A X X A X X X P L X V X X V L X X F L X X L X X A J K X S P K

TF₇₂ 79 D V A S A A K L L P L Q V R E V L V Q F L M S L A D A I K X S P K

TF₆₃ 80 D V A A M A R Q F P L K V K R V L I R F L D E L S E A L K H



e

Consensus: MXXXXXXXXXXKXXLXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXLXXAXXXLXXXVXXXXXXXXXXXXJXALLHDAEDMLEILIQ
 smORF₇₂: 1 MHQNKNDKRSQALLVL-----15-----16 MLQLKDAREKLRNQVAKDHHINDDLALLHDAEDMLEILIQ 21 31 41 56
 smORF₆₃: 1 MKNLEIKKIIGDLEEMAAAAEPGVVITEEMSISCSVTLLKAVDELRKRVGKDISTQTAIV 10 20 30 40 50 60

Extended Data Figure 3. Sequence alignment of phage 72 and phage 63 TF proteins and characterization of the phage 63 TF₆₃-smORF₆₃ module in *Vibrio* 1F-97.

(a) Protein sequence alignment (ClustalW) showing TF₇₂ and TF₆₃. Black boxes show identical residues. The X symbols in the consensus sequence designate different residues. Stars below residues indicate the 5 amino acids in each protein that confer promoter specificity and that have been exchanged in TF_{72-5x} and TF_{63-5x} (see Figure 2b).

(b) Detection of phage 63-specific particles in viral preparations of culture fluids from *Vibrio* 1F-97 carrying aTc-inducible smORF₆₃ that were grown in medium with ciprofloxacin or aTc to induce SOS and smORF₆₃ production, respectively. Relative viral load is the amount of *cl*₆₃ in the induced samples relative to an uninduced sample as judged by qPCR.

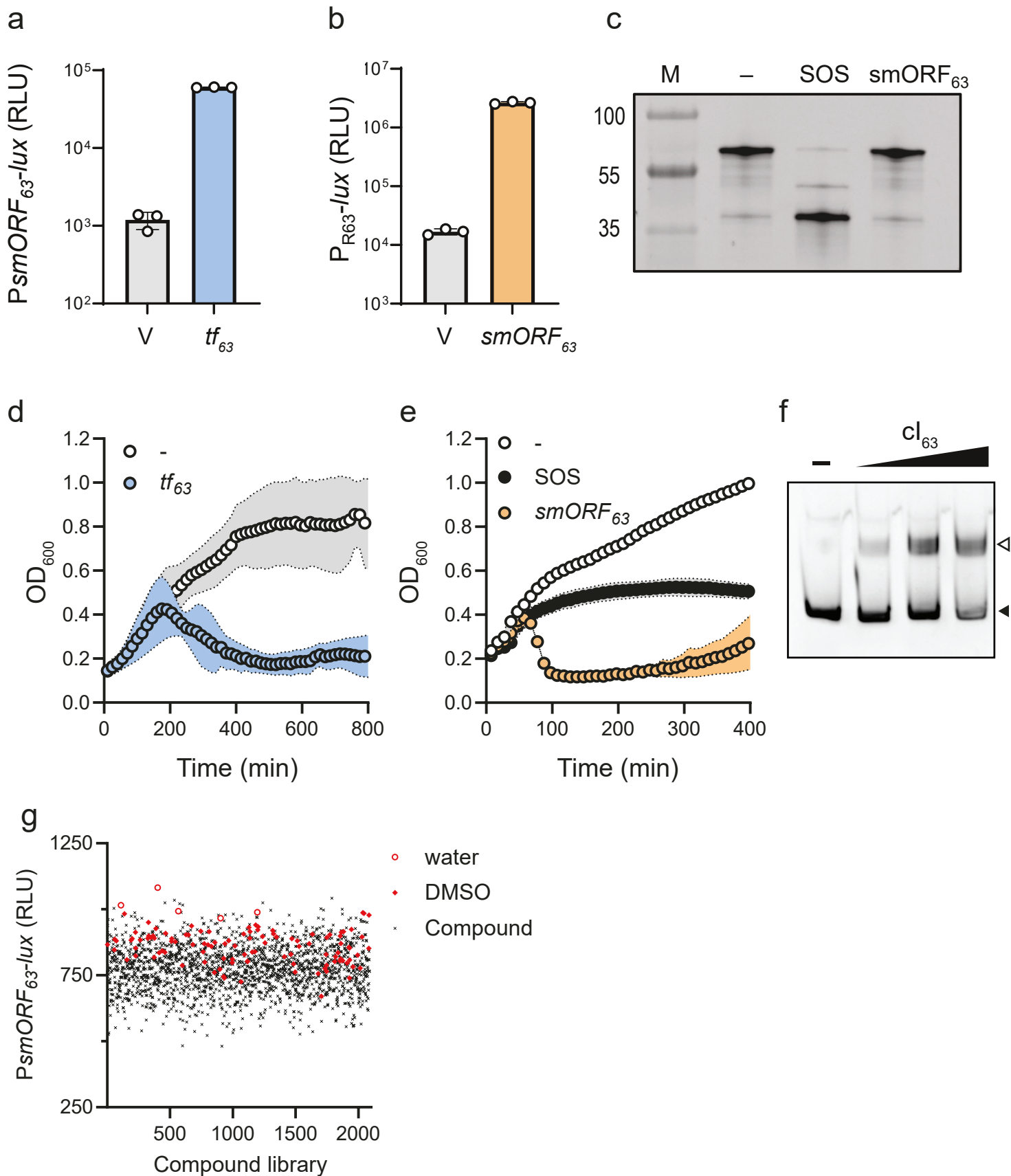
(c) Growth of *Vibrio* 1F-97 carrying a plasmid containing the intervening gene between *repA* and *telN* from phage 63 under an aTc-inducible promoter (black), a plasmid carrying aTc-inducible smORF₇₂ (orange), or no plasmid (white). All media contained aTc.

(d) Organization of genes encoded between *repA* and *telN* in phage 63 and phage 72.

(e) Protein sequence alignment as in (a), for smORF₇₂ and smORF₆₃. Colors and symbols as in (a).

Data are represented as means ± std with *n*=3 biological replicates (c) and as means ± std with *n*=3 biological replicates and *n*=4 technical replicates (b). aTc; 50 ng mL⁻¹ (b, c), ciprofloxacin; 500 ng mL⁻¹ (b).

Extended Data Figure 4



Extended Data Figure 4. TF₆₃ activates transcription of *smORF*₆₃ and *smORF*₆₃ non-proteolytically inhibits *cl*₆₃, driving host-cell lysis.

(a) *PsmORF*₆₃-*lux* expression from *E. coli* carrying an empty vector (V) or aTc-inducible *tf*₆₃ in medium containing aTc.

(b) *P*_{R63}-*lux* expression in *E. coli* carrying an empty vector (V) or aTc-inducible *smORF*₆₃ in medium containing aTc. The *P*_{R63}-*lux* plasmid carries *cl*₆₃, which natively represses reporter expression.

(c) SDS-PAGE in-gel labeling of the phage 63 repressor (HALO-*cl*₆₃) produced in *E. coli* carrying aTc-inducible *smORF*₆₃. The treatments -, SOS, and *smORF*₆₃ refer to water, ciprofloxacin, and aTc, respectively. M as in Figure 1b.

(d) Growth of *Vibrio* 1F-97 carrying aTc-inducible *tf*₆₃ in medium lacking or containing aTc (white and blue, respectively).

(e) Growth of *Vibrio* 1F-97 carrying aTc-inducible *smORF*₆₃ in medium containing aTc (designated *smORF*₆₃, orange), ciprofloxacin (designated SOS, black), or water (designated -, white).

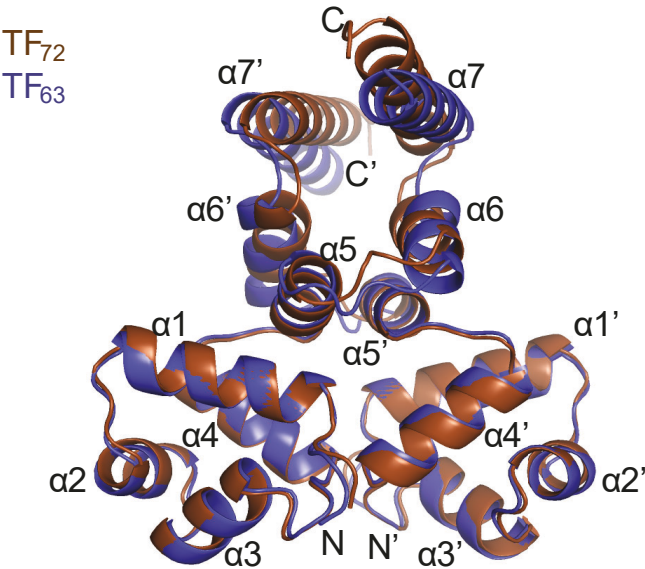
(f) EMSA showing binding of *cl*₆₃ protein to *P*_{R63} DNA. Approximately 8 nM of *P*_{R63} DNA was combined with 800, 1600, or 3200 nM of *cl*₆₃ protein. The no protein control lane is designated with a minus sign. Locations of the unshifted and shifted probe are indicated with black and white arrows, respectively.

(g) Relative *PsmORF*₆₃-*lux* expression in *Vibrio* 1F-97 cultured in Biolog microarray plates and with a curated library of antibiotics. The red circles and diamonds represent water and DMSO vehicle controls, respectively. The black x symbols show the results for the different compounds tested. Assessment of reporter function is provided in (a).

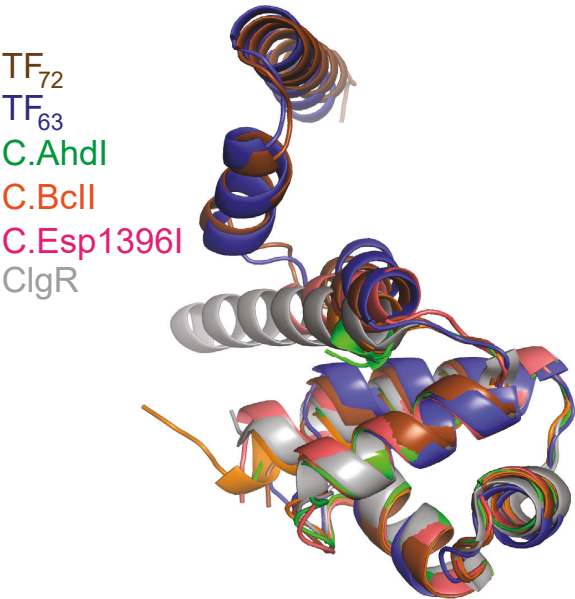
RLU as in Figure 1d (a, b, g). Data are represented as means ± std with *n*=3 biological replicates (a, b, d, e), as a single reading (g), and as a single representative image (f). aTc; 50 ng mL⁻¹ (a, b, c, d, e), ciprofloxacin; 500 ng mL⁻¹ (c, e).

Extended Data Figure 5

a



b

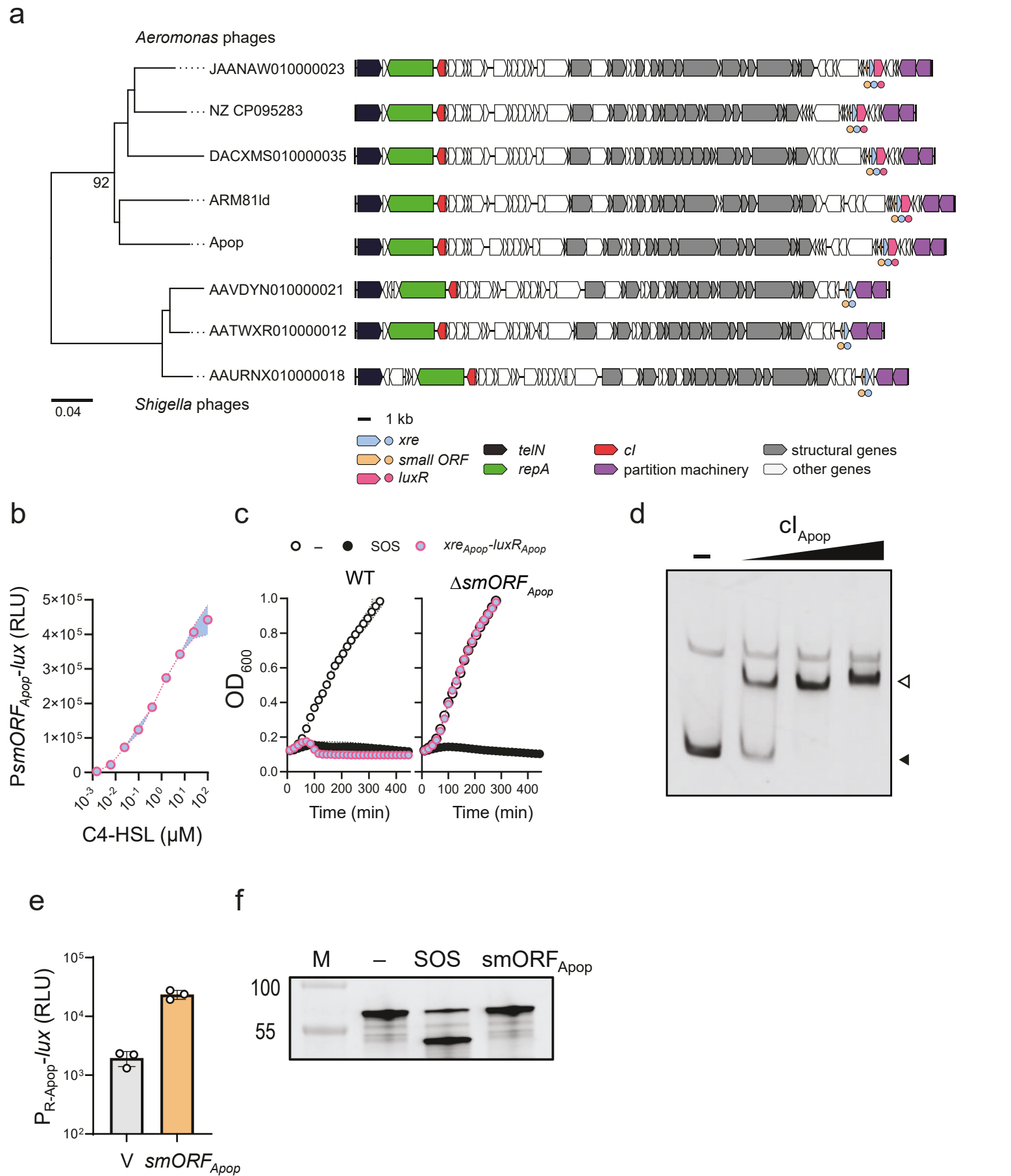


Extended Data Figure 5. Phage TFs are predicted to be structurally similar to Restriction-Modification system controller proteins and to ClgR.

(a) AlphaFold predictions for TF₇₂ (brown) and TF₆₃ (blue) shown as superimposed homodimers. Secondary structural alpha helix elements are labeled. N and C termini are labeled N, N' and C, and C', respectively.

(b) Structural alignment of TF₇₂ (brown) and TF₆₃ (blue) as monomers with the highest scoring homologs: C.AhdI (green), C.BclI (orange), C.Esp1396I (pink), and ClgR (gray) (PDB ID: 1Y7Y, 2B5A, 3G5G, and 5WOQ, respectively).

Extended Data Figure 6



Extended Data Figure 6. Comparison of linear plasmid-like phages in *Shigella* and *Aeromonas* reveals similar locations of genes encoding TF-smORF regulatory modules and HSL-quorum-sensing-receptor TF-smORF modules that control lysis.

(a) Phylogenetic tree (left) of 5 *Aeromonas* phages encoding *xre-luxR* genes and 3 *Shigella* phages encoding *xre*. The genome organization for each phage is depicted at the right. Genes are colored by annotation as noted in the key. Circles denote the locations of relevant features that are common among the 8 phages (*xre* (blue) and *smORF* (orange)) or exclusive to the 5 *Aeromonas* phages (*luxR* (pink)). *Shigella* phages are labeled with their corresponding NCBI accession numbers. Numbers above branches are pseudo-bootstrap support values from 100 replications.

(b) *PsmORF_{Apop}-lux* expression from *E. coli* carrying aTc-inducible *xre_{Apop}-luxR_{Apop}* in medium with aTc and the indicated concentrations of C4-HSL.

(c) Growth of *A. popoffii* carrying WT Apop (left) or $\Delta smORF_{Apop}$ Apop (right), each harboring aTc-inducible *xre_{Apop}-luxR_{Apop}* and grown in medium containing 5 ng mL⁻¹ aTc (blue/pink), 1 μ g mL⁻¹ ciprofloxacin (black), or water (white). All media contained 10 μ M C4-HSL.

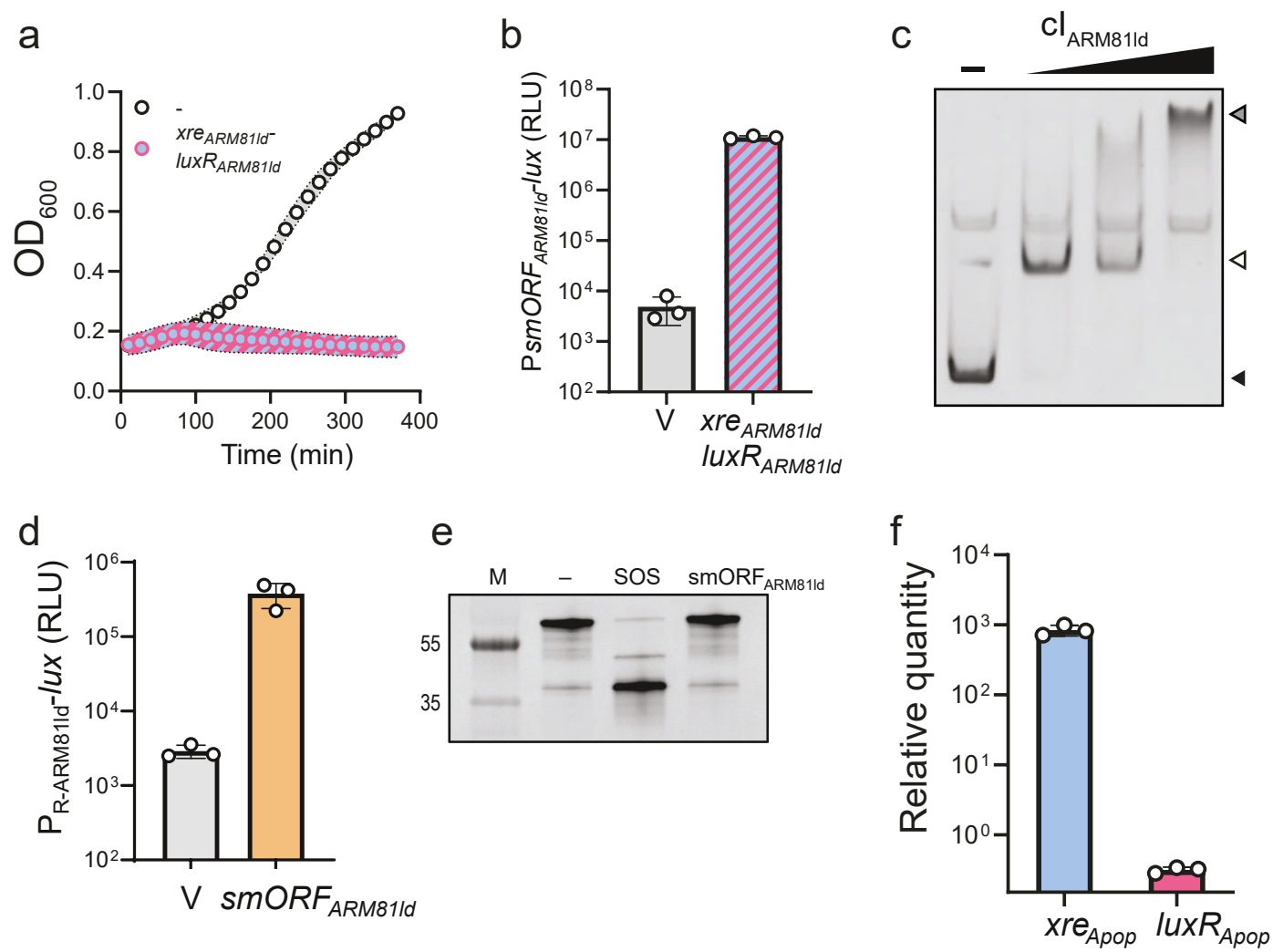
(d) EMSA showing binding of *cl_{Apop}* protein to *P_{R-Apop}* DNA. Approximately 10 nM of *P_{R-Apop}* DNA was combined with 200, 400, or 800 nM of *cl_{Apop}* protein. Locations of the unshifted and shifted probe are indicated with black and white arrows, respectively.

(e) *P_{R-Apop}-lux* expression from *E. coli* carrying an empty vector (V) or aTc-inducible *smORF_{Apop}* in medium containing aTc. The *P_{R-Apop}-lux* plasmid carries two copies of *cl_{Apop}* (see Methods) for native repression of reporter expression.

(f) SDS-PAGE in-gel labeling of the Apop repressor (HALO-*cl_{Apop}*) produced in *E. coli* carrying aTc-inducible *smORF_{Apop}*. The treatments -, SOS, and *smORF_{Apop}* refer to water, ciprofloxacin, and aTc, respectively. M as in Figure 1b.

Data are represented as means $\pm$ std with *n*=3 biological replicates (b, c, e) and as a single representative image (d, f). RLU as in Figure 1d (b, e). aTc; 50 ng mL⁻¹ (b, e, f), ciprofloxacin; 500 ng mL⁻¹ (f).

Extended Data Figure 7



Extended Data Figure 7. The TF-smORF modules of the Apop and ARM81ld phages function analogously, and both XRE and LuxR are required to bind and activate the cognate PsmORF.

(a) Growth of *Aeromonas* sp. ARM81 harboring aTc-inducible $xre_{ARM81ld}$ - $luxR_{ARM81ld}$ in the presence and absence of aTc (blue/pink and white, respectively). All media contained C4-HSL.

(b) $PsmORF_{ARM81ld}$ - lux expression in *E. coli* carrying an empty vector (V) or aTc-inducible $xre_{ARM81ld}$ - $luxR_{ARM81ld}$ in medium containing aTc and C4-HSL.

(c) EMSA showing binding of $cl_{ARM81ld}$ protein to $P_{R-ARM81ld}$ DNA. Approximately 10 nM of $P_{R-ARM81ld}$ DNA was combined with 200, 400, or 800 nM of $cl_{ARM81ld}$ protein. The no protein control lane is designated with a minus sign. Locations of the unshifted, shifted, and supershifted probe are indicated with black, white, and gray arrows, respectively.

(d) $P_{R-ARM81ld}$ - lux expression from *E. coli* carrying an empty vector (V) or aTc-inducible $smORF_{ARM81ld}$ in medium containing aTc. The $P_{R-ARM81ld}$ - lux plasmid carries two copies of $cl_{ARM81ld}$ (see Methods) for native repression of reporter expression.

(e) SDS-PAGE in-gel labeling of the ARM81ld repressor (HALO- $cl_{ARM81ld}$) produced in *E. coli* carrying aTc-inducible $smORF_{ARM81ld}$. The treatments -, SOS, and $smORF_{ARM81ld}$ refer to water, ciprofloxacin, and aTc, respectively. M as in Figure 1b.

(f) Relative transcript levels of the xre_{Apop} - $luxR_{Apop}$ operon from the Apop genome following plasmid expression of either aTc-inducible xre_{Apop} or $luxR_{Apop}$ in *A. popoffii*. All media contained C4-HSL and aTc. Primer pairs specific to the intergenic region in the xre_{Apop} - $luxR_{Apop}$ locus but absent from the aTc-inducible xre_{Apop} and $luxR_{Apop}$ plasmids were used to measure native xre_{Apop} - $luxR_{Apop}$ expression (see Methods). Relative transcript levels are the amount of xre_{Apop} - $luxR_{Apop}$ DNA relative to the amount of $rpoB$ DNA, normalized to the sample overexpressing $luxR_{Apop}$. Data are represented as means $\pm$ std with $n=3$ biological replicates (a, b, d), as means $\pm$ std with $n=3$ biological replicates and $n=4$ technical replicates (f), and as a single representative image (c, e). RLU as in Figure 1d (b, d). aTc; 50 ng mL⁻¹ (b, d, e), ciprofloxacin; 500 ng mL⁻¹ (e).

METHODS

Sequence Retrieval, RepA/TelN and ParB-associated loci

To identify examples of phage genomes with convergently-oriented *telN* and *repA* genes, we examined sequences from the following databases: NCBI nt, IMG/VR v3 (specifically, the file 'IMGVR_all_nucleotides.fna'),¹ Cenote Human Virome Database v1.1 (CHVD_clustered_mash99_v1.fna),² Global Ocean Virome 2 database (GOV2_viral_populations_larger_than_5Kb_or_circular.fna),³ the Gut Phage Database (GPD_sequences.fna),⁴ a curated set of linear plasmid-phages (retrieved via NCBI accession numbers in Table S4 of Pfeifer *et al*),⁵ and the Metagenomic Gut Virus Catalog (mgv_contigs.fna).⁶ In February of 2022, most databases were searched using both tBLASTn and profile-HMM-based search strategies (the exception being NCBI nt, which was too large to annotate anew using profile HMMs). A manual examination of established lysis-control loci⁷ revealed that the associated *telN* gene always encoded a protein that matched closely to Pfam profile PF16684, whereas the following set of seven families encompassed the diversity of observed *repA* sequences: PF13362, PF02399, PF08707, PF10661, PF02502, and PF13604. Thus, we extracted these seven families from Pfam (v34)⁸ into a custom HMM database. We then used the gene finder MetaGeneMark⁹ to predict open reading frames (ORFs) from the aforementioned nucleotide files using default parameters. The resulting proteins were used in a profile HMM search with HMMER3⁹ against the *repA/telN* custom database and nucleotide sequences were extracted that contained both *telN* and *repA* genes, only if they were convergently oriented and within 10 Kb of one another. To complement this search strategy, we also used the predicted TelN and RepA proteins from vibriophage VP882 in a tBLASTn search against these nucleotide databases (accession numbers YP_001039865 and YP_001039868, respectively). We retained only the hits with e-values better than 0.001 which also covered > 50% of the query sequence. For consistency, we re-annotated all retrieved sequences using a common

method, as follows. We used the gene-finder MetaGeneMark¹⁰ to predict open reading frames (ORFs) using default parameters. We next used their amino acid sequences in a profile HMM search with HMMER3⁹ against TIGRFAM¹¹ and Pfam⁸ profile HMM databases. The highest scoring profile was used to annotate each ORF. As above, we further refined our database by considering only contigs with convergently oriented *telN* and *repA* genes within 10 Kb of one another. The HMM and tBLASTn-based search strategies produced highly redundant (but not identical) sequence sets, as the same databases were used and many identical sequences are listed across multiple databases. Thus, we dereplicated our combined sequence files using cd-hit-est with the following parameters: '-c 1.0 -aS 1.0 -g 1 -d 0'. We manually examined the DNA sequence located between each *telN* and *repA* gene from this dereplicated set, extracting the intervening nucleotides to a second locus-specific dataset. We dereplicated these sequences as described above to produce the final set of 274 loci referenced in the text and detailed in Supplementary Table 1. The NCBI nucleotide accession that encodes phage 63 (NZ_AIDA02000063) is dated 20-March-2022, after our initial February search. The sequence for phage 72 appears more than once in the NCBI nucleotide database (NZ_AIDA02000072, dated 20-March-2022, and KP795532, dated 18-Nov-2019). Presumably, these similar entries result from separate sequencing analyses of identical (or similar) bacteria. For these reasons, our February BLAST search only initially revealed phage 72, while phage 63 was discovered subsequently.

To investigate phage genomes for potential TF₆₃-associated marker genes, we first performed a preliminary BLASTp search against the NCBI nr/nt database using TF₆₃ as a query (accession number WP_016786069). We filtered the top 5,000 hits from this search for those with over 35% amino acid identity across 70% of the query sequence and retrieved the corresponding nucleotide file in NCBI, retaining the DNA +/- 50 Kb from the gene boundaries. This analysis produced 744 sequences which were then filtered for those of predicted phage origin using the phage prediction

tool VIBRANT¹² (v1.2) with default parameters. VIBRANT indicated that 200 sequences in this set were putatively phage-derived. Some of these putative phage TF₆₃ sequences aligned poorly to our original query so a second tBlastn search was performed using the original TF₆₃ protein. From the 200 putative phage genomes, only those with tBlastn hits that had an e-value better than e^{-10} were retained. Finally, we removed one short sequence (<5 Kb) and dereplicated the remaining phage sequences at 95% nucleotide identity with cd-hit-est using the following parameters: '-c 0.95 -aS 0.95 -g 1 -d 0'. This produced 117 sequences which were then uniformly annotated as described earlier and manually categorized into the genomic context groups described in Supplementary Table 2. The phage prediction tool VIBRANT was also used to assign a likely phage origin to the sequences listed in Supplementary Table 1, as described above.

To identify examples of putative lysis-control loci associated with the *parB* gene, we again performed a BLASTp search against the NCBI nr/nt database, this time using the ParB protein from phage 63 as a query (accession number WP_016786072). From the top 10,000 proteins revealed by this search, we retrieved the corresponding nucleotide file in NCBI and examined the locus surrounding *parB* by downloading DNA +/- 50 Kb from the gene boundaries (this strategy retrieved the full contig sequence for many of the associated nucleotide files). To identify *parB* genes associated with the phage structural locus present in phage 63, we used tBLASTn to query these sequences with the major capsid protein from this phage (accession number WP_016786053), which is one of the most conserved phage protein folds.¹³ We filtered for hits with over 25% amino acid identity across 70% of the query sequence, revealing 121 putative *parB* loci present on 118 unique sequences. These 121 loci were dereplicated as described for *telN/repA* and filtered for the presence of a full-length *parB* gene and a predicted transcription factor within 10 Kb of *parB*. This analysis produced 56 sequences, which are referenced in Supplementary Table 3.

Phylogenetic analysis of TelN and ParB

From a manual examination of the retrieved RepA- and TelN-encoding sequences, we observed that all TelN proteins shared the same profile HMM as their top-scoring annotation (PF16684.8). This pattern stood in contrast to predicted RepA-encoding sequences, which hit best to one of six different protein families. The consistent TelN annotation suggested to us that this protein may be conserved despite extensive sequence divergence between the phage genomes considered, and thus, it provided a good phylogenetic bellwether for this locus. Indeed, the set of non-redundant, full-length TelN proteins from these datasets aligned well using MUSCLE with default parameters.¹⁴ A preliminary phylogenetic tree was generated from these sequences using the Geneious Tree Builder per the UPGMA clustering method and with a Jukes-Cantor distance model. From this tree, a subset of 34 sequences was selected to best represent the full TelN phylogenetic tree and the genetic diversity encoded by its associated gene neighborhoods. Figure 1a depicts a phylogenetic tree generated from these 34 representative sequences, which was produced using PhyML with the LG substitution model and bootstrapped 100 times. A similar procedure was used to produce the ParB phylogeny depicted in Figure 3a. In this case, a preliminary phylogenetic tree was produced from the set of 56 ParB proteins in the full dataset using the Geneious Tree Builder per the UPGMA clustering method and with a Jukes-Cantor distance model. From this tree, a subset of 28 sequences was selected to best represent the full ParB phylogenetic tree and the genetic diversity encoded by its associated gene neighborhoods. This final ParB tree was produced using PhyML with parameters identical to those used for TelN.

Prediction of autoproteolytic cl proteins

All 271 predicted cl proteins were used to perform a batch search of the NCBI Conserved Domain (CD) database. The output of this search included conserved domains and other protein features for each cl protein queried. Among these features were predicted catalytic sites that corresponded to known catalytic residues in canonical autoproteolytic cl proteins.¹⁵ Thus, we used this feature

table to define the 61 predicted autoproteolytic cl proteins and labeled the remaining 210 proteins as putatively non-autoproteolytic. The clustering analysis in Supplementary Table 1 (summary statistics tab) was performed using the DNA sequences between *repA* and *tefN* genes and grouped sequences that shared 80% nucleotide identity over 95% the length of the shorter sequence in the same cluster. We used the program 'cd-hit-est' with the following parameters: -c 0.8 -aS 0.95 -g 1 -d 0.

Phylogenetic analysis of linear plasmid-like phages encoding *par*-associated *xre* modules in *Shigella* and *Aeromonas*

NCBI BLASTp of the DNA sequence encoding XRE_{Apop} was used as a representative of the three *Aeromonas* phages identified in the *par*-associated analysis (detailed above) retrieved three predicted linear plasmid-like phages in *Shigella*. For each phage genome, genes were called using Prodigal 2.6.3,¹⁶ and gene diagrams were constructed using custom python scripts. Genes were annotated using Prokka 1.11,¹⁷ and annotations were supplemented with NCBI BLASTp searches by hand.¹⁸ The phage genome tree (Extended Data Figure 3a) was constructed using VICTOR,¹⁹ yielding an average support of 78%. The numbers above branches are pseudo-bootstrap support values from 100 replications.

Bacterial strains and growth conditions

E. coli and *Aeromonads* were grown with aeration in Luria-Bertani (LB-Miller, BD-Difco) broth. *Vibrio* strains were grown in LB with 3% NaCl. All strains were grown at 30° C. Strains used in the study are listed in Supplementary Table 4. Unless otherwise noted, antibiotics, were used at: 100 µg mL⁻¹ ampicillin (Amp, Sigma), 50 µg mL⁻¹ kanamycin (Kan, GoldBio), and 5 µg mL⁻¹ chloramphenicol (Cm, Sigma). Inducers were used as follows: *E. coli*: 200 µM isopropyl beta-D-1-thiogalactopyranoside (IPTG, GoldBio), and 50 ng mL⁻¹ anhydrotetracycline (aTc, Clontech). *Vibrios*: 500 ng mL⁻¹ ciprofloxacin (Sigma) and 50 ng mL⁻¹ aTc. *Aeromonads*: 1 µg mL⁻¹

ciprofloxacin and 5 ng mL⁻¹ aTc. C4-HSL was supplied at a final concentration of 10 µM except for the experiment shown in Figure 3b, in which it was administered at the indicated concentrations.

Cloning techniques

All primers and dsDNA (gene blocks) used for plasmid construction, qPCR, and EMSAs listed in Supplementary Table 5, were obtained from Integrated DNA Technologies or Twist Bioscience. Gibson assembly, intramolecular reclosure, and traditional cloning methods were employed for all cloning. PCR with iProof was used to generate insert and backbone DNA. Gibson assembly relied on the HiFi DNA assembly mix (NEB). The Apop- and ARM81Id-based P_R-*lux* reporter constructs in *E. coli* (JSS-3346k and JSS-3348k, respectively) required the addition of a second copy of the *cl* repressor under its native promoter in the plasmid for function (see Supplementary Table 5). All enzymes used in cloning were obtained from NEB. Plasmids used in this study are listed in Supplementary Table 6. Transfer of plasmids into *Vibrio* 1F-97, *A. popoffii*, and *Aeromonas* sp. ARM81 was carried out by conjugation followed by selective plating on TCBS agar supplemented with Kan, LB plates supplemented with Amp and Kan, and LB plates supplemented with Kan and Cm, respectively.

Growth, lysis, and reporter assays

Overnight cultures were back-diluted 1:100 with fresh medium with appropriate antibiotics prior to being dispensed (200 µL) into 96 well plates (Corning Costar 3904). Cells were grown in the plates for 90 min before ciprofloxacin, aTc, or C4-HSL were added as specified. Wells that did not receive treatment received an equal volume of water or DMSO. A BioTek Synergy Neo2 Multi-Mode reader was used to measure OD₆₀₀ and bioluminescence. Relative light units (RLU) were calculated by dividing the bioluminescence readings by the OD₆₀₀ at that time.

RT-qPCR

Overnight cultures were back-diluted 1:100 and grown for 90 min prior to administration of aTc or ciprofloxacin at the indicated concentrations. In Extended Data Figures 1d and 7f, cells were collected at T = 0 and 15 min, and in Extended Data Figure 1h, at T = 0 and 90 min following induction. Harvested cells were treated with RNAProtect Bacteria Reagent (Qiagen) according to the supplier's protocol. Total RNA was isolated from cultures using the RNeasy Mini Kit (Qiagen). RNA samples were treated with DNase using the TURBO DNA-free Kit (Thermo). cDNA was prepared as described using SuperScriptIII Reverse Transcriptase (Thermo). SYBR Green mix (Quanta) and Applied Biosystems QuantStudio 6 Flex Real-Time PCR detection system (Thermo) were used for real-time PCR. Each cDNA sample was amplified in technical quadruplicate and data were analyzed by a comparative CT method which the indicated target gene was normalized to an internal bacterial control gene (*rpoB*).

qPCR and viral preparation

Viral preparations consisted of non-chromosomal DNA (RQ1, RNase-Free DNase, Promega) prepared from 1 mL of cells of the indicated strains. Overnight cultures were back-diluted 1:100 and grown for 90 min before being divided into 3 equal volumes and exposed to treatments as specified. Cultures were grown for an additional 5 h prior to collection of cell-free culture fluids (Corning SpinX). qPCR reactions were performed as described above for RT-qPCR reactions. 1 μ L of purified non-chromosomal DNA was used for each qPCR reaction. Data were analyzed by normalizing the CT values of samples treated with ciprofloxacin or aTc to the CT values of samples treated with water using a primer set to the indicated phage. Viral preparations for whole genome sequencing (SeqCenter) were prepared exactly as described for qPCR, except by column purification (Phage DNA Isolation Kit, Norgen Biotek).

Small molecule-screening for *PsmORF₇₂*- and *PsmORF₆₃-lux* inducers in *Vibrio* 1F-97

Overnight cultures of *Vibrio* 1F-97 strains harboring a plasmid encoding either *PsmORF₇₂-lux* or *PsmORF₆₃-lux* were back-diluted 1:100 and grown for 90 min prior to being dispensed into 23 x 96-well Biolog plates (PM3b-PM25) and 3 x 96 well plates of curated compounds from the Seyedsayamdost Group (Princeton). Plates were incubated at 30° C overnight, prior to measurement of OD₆₀₀ and bioluminescence. Biolog conditions or library compounds that inhibited growth were retested in a *Vibrio* 1F-97 strain harboring a reporter for phage lambda induction (*cl-P_R-lux*).²⁰ The lambda reporter is SOS-responsive but not responsive to a TF-smORF module. The goal was to determine whether any condition/compound was a general phage inducer.

Single molecule RNA-fluorescent in situ hybridization (smRNA-FISH)

Custom Stellaris RNA FISH Probes were designed against sets of early lytic genes in each phage under study using the Stellaris RNA FISH Probe Designer version 4.2 (LGC, Biosearch Technologies, Petaluma, CA; see Supplementary Table 7). Probes were labeled with Quasar 670 (phage 63 and phage ARM81ld) or CAL Flour Red 590 (phage 72 and phage ARM81mr) (Biosearch Technologies). smFISH was performed as described previously,²¹ with minor modifications. Briefly, cells were fixed with 1 mL cold 1x PBS and 3.7% formaldehyde and mixed at room temperature for 30 min. Samples were subjected to centrifugation at 400 x g for 8 min and the clarified supernatant discarded. Cells in the pellets were washed twice with 1x PBS and subjected to centrifugation at 600 x g for 3.5 min following each wash. Cells in the pellets were resuspended in 300 µL water and permeabilized by addition of 700 µL 100% EtOH with mixing at room temperature for 1 h. Samples were subjected to centrifugation at 600 x g for 7 min and the supernatant discarded. Cells in the pellets were resuspended in 1 mL Stellaris RNA FISH Wash Buffer A (Biosearch Technologies) containing 10% formamide, mixed for 5 min at room temperature, and the cells were collected by centrifugation at 600 x g for 7 min. Cells in the pellets were resuspended in 50 µL Stellaris RNA FISH Hybridisation Buffer containing 10% formamide

and 4 μ L of each probe stock (12.5 μ M) and incubated overnight at 37° C. A 10 μ L aliquot of the hybridization reaction was added to 200 μ L Wash Buffer A followed by centrifugation at 600 x g for 3.5 min. Samples were resuspended in 200 μ L Wash Buffer A, incubated for 30 min at 37° C, and subjected to centrifugation at 600 x g for 3.5 min. This step was repeated. Cells were stained with 50 μ g/mL DAPI in Wash Buffer A for 20 min at 37° C, washed with 200 μ L Stellaris RNA FISH Wash Buffer B, and resuspended in 10 μ L 1x PBS. 1 μ L aliquots of these cell samples were dispensed into No. 1.5 glass coverslip bottomed 24-well plates (MatTek, Ashland, MA, USA), along with 30 μ L VectaShield Mounting Medium and the samples were covered with agarose pads. Imaging was performed with a Nikon Eclipse Ti2 inverted microscope equipped with a Yokogawa CSU-W1 SoRa confocal scanning unit. Samples were imaged with a CFI Apochromat TIRF 60x oil objective lens (Nikon, 1.49 numerical aperture) with excitation wavelengths of 405, 561, and 640 and 0.4 μ m (for *Aeromonas* phages) or 1 μ m (for *Vibrio* phages) z-steps. Images were captured through a 2.8x SoRa magnifier. Images were processed using Nikon NIS-Elements Denoise.ai software.

Cells in the images were segmented in Fiji software from maximum z-projections of the DAPI channel. Groups of cells that could not be accurately resolved were excluded from downstream analysis, and coordinates of remaining cells exported. smFISH data were analyzed using custom python scripts. Briefly, images were convolved with a Gaussian to remove noise. Spots were next detected as local maxima with intensities greater than a threshold set based on a negative control image (negative controls for each probe set are as follows: phage 72, *smORF*₆₃ induced; phage 63, *smORF*₇₂ induced; phage ARM81mr, *smORF*_{ARM81ld} induced; phage ARM81ld, no induction). Each spot was fitted with a 3D Gaussian function to determine the integrated, background-subtracted spot intensity. In instances with multiple spots residing in close proximity, a 3D multi-Gaussian fit was performed. Spots were assigned to cells and the summed intensity from all spots in a cell were reported. Cells with total intensities ≤ 1 were assigned a pseudovalue of 1.

Supplementary Table 8 provides summary data for *Vibrio* 1F-97 phage 72 and phage 63 and Supplementary Table 9 provides summary data for *Aeromonas* sp. ARM81 phage Id and phage mr.

***in vitro* HALO-cl repressor cleavage and in-gel HALO detection**

Assessment of cleavage of HALO-cl proteins in response to DNA damage or smORF induction was carried out in *E. coli* according to a previously described method,⁷ with minor modifications. Briefly, overnight cultures of *E. coli* T7Express lysY/I^q carrying the indicated HALO fusion plasmid and cognate, aTc-inducible smORF vector were diluted 1:200 in medium and grown for 2.5 h with shaking. 200 μ M IPTG was added to the cultures before they were divided into 3 equal volumes followed by administration of the relevant treatment as specified. The treated cultures were incubated without shaking for an additional 2.5 h. Cells were collected by centrifugation (16,100 x g for 1 min), resuspended in BugBuster containing 1 μ M HALO-Alexa₆₆₀ (excitation/emission: 663/690 nm). The cleared supernatant, collected after centrifugation of the lysate (16,100 x g for 10 min), was loaded onto a 4-20% SDS-PAGE stain-free gel. Gels were imaged using an ImageQuant LAS 4000 imager under the Cy5 setting for HALO-Alexa₆₆₀ before being exposed to UV-light for 7 min and re-imaged under the EtBr setting for total protein. Exposure times never exceeded 30 sec.

Protein production and purification

Plasmids harboring genes encoding HIS-HALO-cl_{Apop} and HIS-HALO-cl_{ARM81Id} were introduced into *E. coli* BLR21(DE3) (Millipore Sigma) and a plasmid carrying HIS-HALO-cl₆₃ was introduced into *E. coli* BL21(DE3) (Invitrogen). For protein production, the strains were grown overnight at 15 °C with 1 mM IPTG. In all cases, cells were pelleted at 3000 x g for 10 min followed by resuspension in Buffer A (150 mM NaCl, 20 mM Tris pH 7.5, 1 mM TCEP). cOmplete EDTA-free protease inhibitor cocktail tablets (Millipore Sigma) and 10 units of DNase I (Thermo) were added

to each resuspended pellet and the cells were lysed by sonication. The insoluble fractions were separated from the soluble material by centrifugation of the lysates at 26,000 x g for 40 min. The soluble fractions were collected and applied to Ni-NTA Superflow resin (Qiagen). The resin was washed with 5 column volumes (CVs) of Buffer A. Proteins were eluted in 2.5 CVs of Buffer A and a gradient of 100-300 mM imidazole. The HIS-HALO tags were cleaved from the cl_{Apop}, cl_{ARM81ld}, and cl₆₃ proteins by treatment with 1 mg HIS-TEV Plus protease overnight at 4 °C. The samples were re-applied to Ni-NTA Superflow resin and the proteins were captured at ~85-95% purity by washing the resin with 1 CV of Buffer A.

cl₇₂-HALO-HIS protein was produced in *E. coli* BLR21(DE3) by growth at 37 °C for 3 h with 1 mM IPTG. cl₇₂-HALO-HIS protein was purified from cell pellets as described above for the other cl proteins, however, to circumvent insolubility and aggregation, the HALO-HIS tag was not cleaved. Following initial purification on Ni-NTA Superflow resin, the eluate was concentrated and loaded onto a Superdex-200 size exclusion column (GE Healthcare) in Buffer A. cl₇₂-HALO-HIS at ~85% purity was collected.

Electrophoretic Mobility Shift Assay (EMSA)

DNA probes P_{R63}, P_{R72}, P_{R-Apop}, and P_{R-ARM81ld} were generated using plasmids JSS-3131, JSS-3129, JSS-3346k, and JSS-3348k, respectively, and the primers listed in Supplementary Table 5. Each EMSA reaction contained 20 ng of DNA probe (8-14 nM depending on the probe). The concentration of cl_{Apop} and cl_{ARM81ld} protein used in each reaction ranged from 200 nM to 800 nM. Higher concentrations of the cl₆₃ and cl₇₂-HALO-HIS proteins (800 nM to 3200 nM) were required due to their limited solubility. The cl proteins with their respective probes were combined in binding buffer (50 mM NaCl, 20 mM Tris pH 7.5, 1 mM TCEP) and incubated at room temperature for 15 min. The samples were subjected to electrophoresis on a Novex 6% DNA retardation gel

(Thermo) in 1x TBE at 100 V for 45 min. Double-stranded DNA was stained with SYBR Green I Nucleic Acid Gel Stain (Thermo) for 20 min. After washing with 1x TBE, gels were imaged using an ImageQuant LAS 4000 imager under the SYBR Green setting.

AlphaFold and DALI structural prediction of phage TF₇₂ and TF₆₃

AlphaFold2²² and AlphaFold-Multimer²³ were used to predict the structures of TF₇₂ and TF₆₃ from the protein sequences. The predicted structures were uploaded to the DALI server²⁴ for a heuristic PDB search. Structural predictions and homologs predicted by the DALI server were aligned and visualized with PyMOL.²⁵

Quantitation and statistical analyses

Software used to collect and analyze data generated in this study consisted of: GraphPad Prism 9 for analysis of growth and reporter-based experiments; Gen5 for collection of growth and reporter-based data; Geneious Prime 2020 and SnapGene v6 for analysis of publicly available data and primer design; QuantStudio for qPCR collection; LASX for acquisition of confocal micrographs; and FIJI for image analyses. Data are presented as the means $\pm$ std. The number of technical and independent biological replicates for each experiment are indicated in the figure legends.

Data and software availability

Data presented in each panel of this study is available in Supplementary Table 10. Unprocessed gels and micrographs from this study are deposited on Zenodo (doi: 10.5281/zenodo.7083051). Other experimental data that support the findings of this study will be provided without restriction by request from the corresponding author.

METHODS REFERENCES

1. Roux, S. *et al.* IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. *Nucleic Acids Res.* **49**, D764–D775 (2021).
2. Tisza, M. J. & Buck, C. B. A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. *Proc. Natl. Acad. Sci.* **118**, e2023202118 (2021).
3. Gregory, A. C. *et al.* Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* **177**, 1109–1123.e14 (2019).
4. Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D. Massive expansion of human gut bacteriophage diversity. *Cell* **184**, 1098–1109.e9 (2021).
5. Pfeifer, E., Moura de Sousa, J. A., Touchon, M. & Rocha, E. P. C. Bacteria have numerous distinctive groups of phage–plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic Acids Res.* **49**, 2655–2673 (2021).
6. Nayfach, S. *et al.* Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat. Microbiol.* **6**, 960–970 (2021).
7. Silpe, J. E. & Bassler, B. L. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. *Cell* **176**, 268–280.e13 (2019).
8. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **28**, 263–266 (2000).
9. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
10. Zhu, W., Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **38**, e132 (2010).
11. Haft, D. H. *et al.* TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res.* **29**, 41–43 (2001).
12. Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**, 90 (2020).
13. Koonin, E. V. *et al.* Global Organization and Proposed Megataxonomy of the Virus World. *Microbiol. Mol. Biol. Rev. MMBR* **84**, e00061–19 (2020).
14. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

- 1155 15. Little, J. W. Autodigestion of *lexA* and phage lambda repressors. *Proc. Natl. Acad. Sci.* **81**,
1156 1375–1379 (1984).
- 1157 16. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site
1158 identification. *BMC Bioinformatics* **11**, 119 (2010).
- 1159 17. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069
1160 (2014).
- 1161 18. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5-9 (2008).
- 1162 19. Meier-Kolthoff, J. P. & Göker, M. VICTOR: genome-based phylogeny and classification of
1163 prokaryotic viruses. *Bioinforma. Oxf. Engl.* **33**, 3396–3404 (2017).
- 1164 20. Silpe, J. E., Wong, J. W. H., Owen, S. V., Baym, M. & Balskus, E. P. The bacterial toxin
1165 colibactin triggers prophage induction. *Nature* **603**, 315–320 (2022).
- 1166 21. Skinner, S. O., Sepúlveda, L. A., Xu, H. & Golding, I. Measuring mRNA copy number in
1167 individual *Escherichia coli* cells using single-molecule fluorescent in situ hybridization. *Nat.*
1168 *Protoc.* **8**, 1100–1113 (2013).
- 1169 22. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–
1170 589 (2021).
- 1171 23. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. 2021.10.04.463034
1172 Preprint at <https://doi.org/10.1101/2021.10.04.463034> (2022).
- 1173 24. Holm, L. Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210–
1174 W215 (2022).
- 1175 25. Delano, W. L. The PyMOL Molecular Graphics System (2002). in (2002).
- 1176

ACKNOWLEDGMENTS

We thank Professor Martin Polz (University of Vienna) for gifting us *Vibrio* 1F-97 and for thoughtful discussions throughout and Professor Mohamed Seyedsayamdost for providing the compound screening library. We thank Chenyi Fei for assistance with smRNA-FISH analyses, and Julie Chen (Broad Institute), Abhishek Biswas (Princeton University), and all members of the Bassler lab for insightful discussions. This work was supported by the Howard Hughes Medical Institute, National Science Foundation grant MCB-1713731, and the National Institutes of Health grant R37GM065859 to B.L.B. J.E.S. is a Howard Hughes Medical Institute Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. O.P.D. was supported by the NIGMS T32GM007388 grant. G.E.J. is a Howard Hughes Medical Institute Fellow of the Damon Runyon Cancer Research Foundation, DRG-2468-22. G.A.B. was supported by NIH Grant F32GM149034. F.A.H. is funded by the Schmidt Science Fellowship. K.J.F. was supported by the Endowed Scholars Program at the University of Texas Southwestern Medical Center, a Searle Scholars award, and NIH grant 1DP2-AI154402. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

AUTHOR CONTRIBUTIONS

J.E.S., O.P.D., F.A.H., K.J.F., and B.L.B. conceptualized the project. J.E.S. and O.P.D. constructed strains. J.E.S., O.P.D., G.E.J., and G.A.B. performed experiments. J.E.S., F.A.H., and K.J.F. performed bioinformatic analyses. J.E.S., O.P.D., G.E.J., G.A.B., F.A.H., K.J.F., and B.L.B. analyzed data. J.E.S., O.P.D., G.E.J., G.A.B., and B.L.B. designed experiments. J.E.S., O.P.D., G.E.J., G.A.B., K.J.F., and B.L.B. wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.

1202

1203 **ADDITIONAL INFORMATION**

1204 Supplementary Information is available for this manuscript. Correspondence and requests for
1205 materials should be addressed to: bbassler@princeton.edu