Activation of programmed cell death and counter-defense functions of phage accessory genes

Sukrit Silas[1], Héloïse Carion[1], Kira S. Makarova[2], Eric Laderman[1], David Sanchez Godinez[3,4], Matthew Johnson[1], Andrea Fossati[5,6,7], Danielle Swaney[5,6,7], Michael Bocek[8], Eugene V. Koonin[2], Joseph Bondy-Denomy[1,6,9]

[1] Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA 94158, USA

[2] National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

[3] Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA 94158, USA

[4] Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA

[5] J. David Gladstone Institutes, San Francisco, 94158, California, USA

[6] Quantitative Biosciences Institute, University of California, San Francisco, San Francisco, CA 94158, USA

[7] Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94158, USA

[8] Twist Biosciences, South San Francisco, CA 94080, USA

[9] Innovative Genomics Institute, Berkeley, CA 94720, USA

Viruses have been evolving host-modifying factors for billions of years. Genomes of bacterial and archaeal viruses are replete with fast-evolving, uncharacterized accessory genes (AGs), most of which likely antagonize host defenses or other viruses[1,2]. Systematic investigation of AGs could uncover a multitude of biological mechanisms involved in virus-host competition, but AG identification in genomic databases remains a challenge. We developed an integrated computational and high-throughput discovery platform to identify AGs in virus genomes and assay their functions in complementary phage infection-dependent and -independent contexts. Our approach showcases how phages interact with the principal layers of antiviral immunity, including cell surface modifications, restriction systems, and abortive infection (Abi) mechanisms, which operate simultaneously in the same host. We discovered multiple Enterobacteriophage AGs associated with counter-defense functions that activate rather than inhibit antiviral immunity in cells, including the surprising finding that anti-restriction AGs elicit programmed cell death (PCD) activity of some restriction-modification (R-M) systems. We propose that counter-defense AGs that trigger PCD create a conundrum for phages whereby keeping the AGs causes PCD but losing them exposes the phage to restriction by bacteria. Strategies employed by viruses to avoid this double jeopardy could be an important factor in virus evolution that remains to be explored.

Introduction

To a large extent, the evolution of life is a story of the virus-host arms race[3]. As hosts evolve elaborate antiviral defense mechanisms, viruses retort with equally versatile counter-defenses. Signatures of this arms-race are evident in prokaryotic genomes, where diverse defense systems tend to be encoded within 'defense islands'[4], and in virus genomes, where hotspots known as accessory regions (ARs) contain clusters of highly variable anti-defense genes[1,2] (Figure 1a). These accessory genes (AGs) are at the forefront of virus-host co-evolution, and many of them are known or predicted to interface with host defenses[2]. Bacteriophage AGs are typically non-essential, undergo frequent horizontal transfer, and diversify rapidly under enhanced evolutionary pressure[1]. The characterized AGs are involved in antagonizing other phages or bacterial defense systems such as restriction-modification (R-M)[5] and CRISPR-Cas[6], but a vast pool of AGs in rapidly growing (meta)genomic sequence databases remains unexplored[7]. Determining the identity and functions of AGs systematically remains an important and unaddressed problem.

Comparative genomic approaches can exploit the highly organized architecture of virus genomes to identify evolutionary hotspots. However, ARs have so far only been characterized in isolated phage-families[1,8-12], and AGs have not been studied at scale. We therefore developed a scalable, high-throughput platform to identify and functionally test AGs from large datasets of phage genomes. Here, we focus on AGs that can inhibit, and in some cases, also activate bacterial immunity. We uncover a multi-layered defense architecture and extensive co-evolution of defense and counter-defense mechanisms, including an R-M system that can trigger PCD in addition to its canonical defense functions and another R-M locus providing decoy immunity previously only known in eukaryotes[13]. On the phage side, we discover proteins that alter host metabolism and modify the cell surface, or inhibit antiviral restriction. Moreover, we identify several counter-defense AGs that trigger abortive infection (Abi) responses and propose that bacteria routinely exploit phage AGs to sense infection.

Results and Discussion

A high throughput screening platform for identifying and testing virus accessory genes

Given that AGs are usually defined within a cluster of closely-related viruses[8-11], AG detection is highly sensitive to the genome clustering strategy. We therefore sought to identify ARs without *a priori* sequence comparison or clustering. All pairwise combinations of all phage genes were evaluated sequentially (Figure 1b) to determine if any pair of conserved genes bounds an AR (Supplementary File 1). We also did not rely on any gene ontology to identify AGs (eg. genes similar to or encoded next to known counter-defense factors). Starting with an initial set of 1706 Enterobacteriophage genomes, 2014 non-redundant ARs containing 10888 putative AGs were identified in 1217 phages. All ARs were then scored and ranked by diversity (See Methods). A sample AR is shown in Figure 1c. A typical AR tallied 11-30 unique AGs cumulatively across multiple related phages (we discarded ARs encountered in <6 genomes). AGs were mostly of ~50-100 aa in length. A typical phage genome contained <5 ARs and <10 AGs, while some phages encompassed many more AGs (Figure 1d).

We manually inspected the 200 highest-scoring ARs and chose multiple (2-5) AGs from 62 of these (see Supplementary File 2 for detailed notes on each AR and rationale for AG selection). We synthesized 200 AGs, of which 54 were from ARs containing known counter-defense factors (Supplementary File 3), whereas 135 AGs were from ARs consisting entirely of genes of unknown function. Eleven phage genes with previously elucidated functions were included as controls (Abc2, ArdA, Arn, Cor, Dmd, Imm, Ip1, Ocr, Ral, SieA, Stp). To identify AG function(s), the genes were barcoded and chromosomally integrated under an IPTG-inducible

promoter in a subset of the *E. coli* Reference (ECOR) strains[14] (one AG per cell) in pools. Fitness was measured in triplicate by next-generation sequencing of AG-specific barcodes upon AG-expression (Figure 1e, Supplementary File 4).

Phage AGs trigger programmed cell death responses in bacteria

Several AGs produced "conditional lethal" phenotypes, imposing a severe fitness burden in a minority of strains, suggestive of PCD induction (Figure 2a). Although most known triggers of abortive defense systems are conserved phage proteins[15-18], we hypothesized that these strains harbor systems that sense variable phage proteins encoded by AGs. Surprisingly, several of these trigger AGs were known anti-restriction factors. Anti-restriction-induced cell death mechanisms might serve as a backup to thwart phage-encoded counter-defense strategies. PCD in these wild strains is triggered by some R-M inhibitors but not others, even discriminating between the two DNA mimics Ocr[19] and ArdA[20] (Supplementary Figure 1a). A novel AG, *orf7* (hereafter, *aop1* for Activator Of PCD) also triggers PCD in ECOR55 and 66 similarly to Ocr and ArdA, but does not inhibit Type I R-M in MG1655 (Supplementary Figure 1b). These observations suggest a diversity of PCD mechanisms.

We transformed the respective wild hosts with the trigger AGs and confirmed that they induce PCD (Figure 2b). We then performed genome-wide transposon suppressor screens to map the loci causing PCD (Figure 2c). Transposon insertions in the causal system would be expected to disable PCD and allow survival of the bacteria despite heterologous expression of a trigger AG. We identified putative PCD systems in all tested hosts, except for ECOR22 with *orf98* and ECOR66 with *orf1*, *orf7*, or *orf169*, where no suppressor mutants could be obtained (Supplementary File 5).

Restriction-modification systems can elicit PCD

Surprisingly, a Type III R-M system in ECOR22 (hereafter, EcoR22I) was required for cell-death in response to Ocr (Figure 2c), the DNA mimic that blocks both Type I and Type III R-M[21]. R-M defenses are ubiquitous in bacteria, and so are anti-restriction factors in phages[22]. In response, bacterial anti-anti-restriction mechanisms such as retrons[23], PARIS[24], and PrrC[25] provide backup defense. Our results suggest that in ECOR22, the frontline (R-M) and backup (PCD) defense functions coalesced into an integrated defense system. To test this hypothesis directly, we reconstituted the PCD and R-M activities of EcoR22I in a model *E. coli* strain.

In its native host, the *mod* and *res* genes of EcoR22I are associated with an SNF2-like helicase related to the ancillary DISARM helicase, *drmD*[26], and all 3 genes are encoded in a P4 satellite prophage hotspot[24]. However, transposon insertions that suppressed Ocr-toxicity in ECOR22 were found only in *mod* and *res* (Figure 2c, Supplementary Figure 2). We overexpressed only the *mod* and *res* genes of EcoR22I in a lab strain of *E. coli* and found that EcoR22I caused PCD upon Ocr induction but not upon induction of ArdA or Aop1, consistent with the behavior of the wild host (Figure 2d). Ocr was not toxic in the lab strain on its own (see Figure 3a). An F54V Ocr mutant, previously reported to evade detection by PARIS[24], also lost the ability to trigger EcoR22I PCD, as did various other Ocr mutants[19] (Supplementary Figure 1c). Because EcoR22I is the only R-M system in the lab strain and in ECOR22, EcoR22I likely detects Ocr directly.

Next, we tested whether the restriction activity of EcoR22I could block phage infection. We constructed an active-site mutant of the PD-(D/E)xK restriction nuclease (res*: D1024A, D1043A double mutant) and compared immunity conferred by wild type and res* EcoR22I against a panel of phages. EcoR22I but not the catalytic site mutant blocked λvir (Figure 2e). Furthermore, λvir passaged through the res* host (and thus modified by the EcoR22I methyltransferase) regained the ability to infect cells carrying wildtype EcoR22I, indicating that EcoR22I indeed restricts only unmodified λvir DNA.

We then tested if the immune (R-M) and PCD functions of EcoR22I could be uncoupled. EcoR22I *mod* and *res* individually did not induce PCD when co-expressed with Ocr, indicating that PCD induction is a property of the complex (Figure 2f). The res* nuclease-dead mutation of EcoR22I that completely abrogated λvir restriction also decreased PCD by 100-fold, although it did not fully abolish Ocr-induced toxicity (Figure 2f). These results show that the EcoR22I PD-(D/E)xK nuclease motif is involved in both R-M and PCD but does not appear to be fully responsible for the latter.

Thus, EcoR22I mounts both a targeted antiviral defense and a PCD response in the presence of certain phage AG products (Figure 2g). Coupling restriction and PCD could allow cells to respond dynamically depending on the progress of the infection[27].

Decoy immunity in bacteria

A second R-M-like system also elicited PCD in the presence of anti-R-M AGs. This locus in ECOR55 contains *hsdS* and *hsdM* genes typical of Type I R-M but lacks the nuclease (*hsdR*), and instead has a tightly linked gene that encodes a 61aa protein. This small gene was identified in our suppressor screens as the PCD-inducing effector activated by Ocr, Aop1, and ArdA (Figure 3a). We named this system Ronin, after the master-less samurai in the 1962 Japanese film, *Harakiri*. Ronin induces cell suicide in ECOR55 when Ocr, ArdA, or Aop1 are expressed (Figure 2b), but only responds to ArdA and Aop1 when cloned into a lab strain (Figure 3b).

Several lines of evidence suggest that the 61aa protein (hereafter, RonA) in Ronin is the PCD effector. The presence of a predicted transmembrane helix suggests RonA might disrupt membrane integrity (Figure 3c). Over-expression of *ronA* alone is lethal, but this effect is suppressed by HsdMS co-expression (Figure 3d). Ronin activity is independent of the methyltransferase function of HsdMS because PCD was not abrogated by an F287G mutation in motif IV (NPPF) of the *hsdM* catalytic domain[28] (Figure 3e). However, removal of *ronA* inactivates Ronin, such that anti-restriction proteins no longer trigger PCD (Figure 3e). The 235-bp non-coding region upstream of *ronA* also contributes an unknown, essential functionality (Supplementary Figure 3). Live cell imaging showed that Ronin activation arrested growth without lysis, consistent with RonA-mediated inner-membrane disruption as a PCD mechanism (Figure 3f). As *ronA* is not lethal when co-expressed with HsdM and HsdS (hereafter, HsdMS), we surmise that its activity is sequestered or suppressed by HsdMS. This model is reminiscent of PrrC activation upon blocking of EcoPrrI restriction by T4-encoded peptide Stp. Similarly to Stp binding EcoPrrI and releasing the Abi anticodon nuclease PrrC[25], an anti-restriction protein could bind Ronin HsdMS to release RonA, the PCD effector.

Examination of the gene neighborhoods of *ronA* homologs showed that they are always linked to *hsdS* and *hsdM* genes and occur in two predominant genetic architectures. Ronin is either linked to an atypical Type I/Type IV composite R-M system or to a BREX system (Figure 3g). The BREX-linked architecture is exclusive to *E. coli*. Most (77%) Ronin-encoding *E. coli* genomes contain no Type I R-M systems anywhere in the genome, and therefore lack *hsdR* genes (Supplementary File 6). The predominantly nuclease-lacking configuration in *E. coli* suggests that PCD is Ronin's primary function. This is reminiscent of decoy immunity (especially common in plants), where a host factor mimics the target of a pathogen effector. The effector then binds the decoy instead of its actual target, and the decoy alerts the host to the presence of the pathogen[13].

Thus, Ronin appears to be a Type I R-M derivative that lost the defense nuclease but gained the ability to sense anti-restriction effectors and trigger PCD. Ronin encodes the canonical target of Ocr/ArdA, the Type I R-M HsdMS complex, and its activation likely depends on these anti-restriction proteins binding HsdMS (Figure 3h). Consistent with this, Ronin could still sense the Ocr F54V mutant that evades PARIS and EcoR22I but retains anti-restriction activity[24]. Surprisingly, however, Ronin also responded to various Ocr mutants that lost their anti-restriction activity (Supplementary Figure 1c). Because it almost always appears linked to an R-M or BREX

system, both of which can be blocked by Ocr[29], Ronin might drive evolution of Ocr-encoding phages to a BREX/R-M susceptible state through the loss of Ocr. Decoy immunity could thus be a powerful host strategy to constrain virus escape from restriction systems in bacteria.

Counter-defense-associated AGs trigger Abi mechanisms encoded in prophages

We identified three prophage-encoded Abi loci that induced PCD in response to counter-defense-associated AGs (Supplementary File 5). First, we found that heterologous *orf116* expression triggered PCD in ECOR15. Transposon-insertions identified the P4 satellite prophage (especially the ~100-bp intergenic region marked by a star in Figure 4a) as the Abi locus. *orf116* is the poorly characterized *abc1* gene of phage P22 that is adjacent to the RecBCD inhibitor *abc2* (Figure 4a)[30]. We cloned the P4 satellite prophage into a plasmid and reconstituted its PCD effect in a lab strain. P4 cannot be induced without a co-resident P2 prophage (which was absent in the lab strain), ruling out prophage-induction as the cause of PCD. Abc1 contains a helix-turn-helix (HTH) domain typical of DNA-binding proteins (Supplementary File 7). Thus, Abc1 might trigger PCD by interfering with P4 transcriptional regulation, perhaps upregulating the lethal gene *kil*.

Second, the gop-beta-cII Abi system is a toxin-antitoxin-like locus located in the P4 hotspot[24,31] with apparent anti-phage activity[24]. Our suppressor screens showed that gop-beta-cII toxicity is activated by *orf184*, which is widely distributed among coliphages and is present in the same AR with the anti-restriction gene *ral* (Figure 4b). We cloned gop-beta-cII into a lab strain and confirmed that *orf184* triggered PCD (Figure 4b).

Finally, a distinct prophage-encoded system in ECOR61 inhibited growth upon expression of *orf24*, which is narrowly distributed and present in the same AR with the anti-restriction gene *ocr* (Figure 4c). We found suppressor transposon-insertions in three prophage genes, including an Icd-like cell-division-inhibitor with a long N-terminal extension (Superfamily cl41269). However, expression of all 6 linked genes (including a DNA primase) was necessary to reconstitute PCD in a lab strain, still yielding a muted effect (Figure 4c) compared to that in the native host (Figure 2b). The Icd-like protein is the likely effector because PCD induction was abrogated upon its removal (Figure 4c). These findings demonstrate that prophages can activate superinfection exclusion mechanisms in response to highly variable phage AGs.

Phage AGs disable host defenses

In the previous sections, we show that phage AGs can activate immune pathways that lead to cell death or dormancy. Next, we wondered how AGs might affect the progress of phage infections in ECOR strains. We layered phage infection onto our AG screens, measuring fitness in the presence or absence of infection by 8 model phages (Figure 1e). Forty-five phage-host combinations were selected where phage replicative titers were highly attenuated compared to growth on a sensitive lab strain (Supplementary Figure 4a). This phenotype suggests that an antiviral mechanism was blocking phage replication, such that expressing an AG that inhibits this defense system could alleviate restriction. Although 8 model phages were used in our infection-based screens, the AGs were taken from hundreds of diverse phage genomes (see Figure 1).

To identify AGs with counter-defense phenotypes, we compared infected samples with uninfected controls and identified several host-phage-AG combinations that were depleted from the infected pools (Figure 5a). Each AG that sensitized any wild strain to infection was then individually validated in that strain against the entire phage set. In these validation experiments, some additional phage-host combinations that were omitted in the initial screen also displayed AG counter-defense phenotypes (Supplementary Figure 4b). In all, increased plaquing of several phages was observed upon expression of 3 known Type I R-M inhibitors (Ocr, Ral, ArdA), 6 novel AGs, and the T4 internal protein 2 (IpII) in a total of 28 phage-host combinations. These AGs likely antagonize host factors that inhibit phage reproduction.

To identify antiviral mechanisms inhibited by AGs, we employed two unbiased approaches: (1) transposon-knockout screens to pinpoint gene disruptions in hosts that phenocopy AG expression (i.e. increased phage infection) (Supplementary File 8) and (2) affinity-purification and mass spectrometry (AP-MS) of AGs to identify binding partners in the wild strains (Supplementary File 9). We performed these assays for every host found to be sensitized to infection by any AG.

AGs antagonize conserved host pathways and barrier defenses

Six novel AGs – *orf48*, *orf63*, *orf92*, *orf116* (*abc1*, see above), *orf126*, and *orf148* – all exhibited counter-defense phenotypes in multiple host strains. Because the broad-spectrum effects of these AGs did not correlate with the variable defense repertoires of these strains (Supplementary Figure 4c), we sought to determine whether they targeted conserved host antiviral pathways.

Genome-wide knockout screens indicated that disrupting the host O-antigen or capsule genes phenocopied the phage-sensitization effects of all 6 broad-spectrum counter-defense AGs with almost all tested phages (Supplementary File 8). Some of the most pronounced phenotypes we observed were that transposon-mediated O-antigen disruption (Figure 5b) or expression of *orf48*, *orf63*, or *orf92* (Figure 5c) both converged on the same effect, namely, enhanced infection by phage T5 (which does not natively encode these AGs). We hypothesized that T5 adsorption was blocked by the wild-type O-antigen structures in these hosts, but heterologous AG expression enabled infection by modifying or removing the O-antigen. We therefore examined LPS preparations from every host where any AG exhibited a counter-defense phenotype. Hosts expressing *orf48*, *orf63*, or *orf92* exhibited subtle downshifts in the O-antigen banding pattern (Figure 5d), as observed previously in some seroconverting lysogens as a means to exclude superinfecting phages[32]. Specifically, the following host-AG combinations showed visible downshifts in O-antigen bands: *orf63* and *orf92* in ECOR1, *orf48* and *orf92* in ECOR3, and *orf48* and *orf63* in ECOR22. These downshifts correlate with the host-AG pairs where T5 infection was especially strongly ($10^{5\text{-}6}$-fold) enhanced (Figure 5c). We reasoned that the AGs might interfere with O-antigen biosynthesis, and tentatively named *orf48*, *orf63* and *orf92*, *gnarl1, 2,* and *3* respectively, after the skin-eating demon in the popular TV series *Buffy the Vampire Slayer*.

All known seroconverting prophages modify O-antigens using enzymes such as glycosyltransferases and acetyltransferases)[32-36]. *Gnarl*-encoding phages in our genomic dataset do not encode any such seroconverting enzymes. Gnarl proteins (40aa – 90aa) are also much smaller than typical enzymes, hinting that these AGs might inhibit O-antigen-modifying enzymes, constituting a distinct mode of host seroconversion. *Gnarl* genes are not restricted to either lytic or lysogenic phages; *gnarl1* and *gnarl3* are present in lytic phages (Supplementary File 10), whereas *gnarl2* is present in the classical temperate phage Mu (gene E6/Mup07). Seroconversion could occur in lysogenic infections to exclude superinfecting phages, or in lytic infections as receptor-masking to prevent newly synthesized virions from unproductively binding lysed cell fragments[37].

To identify the cellular targets of gnarl proteins, we analyzed AP-MS data for all novel counter-defense AGs. Each accessory gene product bound distinct host proteins, although the profiles of *orf116* (*abc1*) and *orf126* were somewhat correlated (Figure 5e). Gene Ontology pathway analysis identified UDP-glucose biosynthesis factors as significantly enriched in the *gnarl3* set, suggesting that *gnarl3* inhibits a step in this pathway. Among these, GalU catalyzes the formation of UDP-glucose[38] to build the LPS outer core[39]; GalF regulates the levels of UDP-glucose[40]; and GalE reversibly converts UDP-glucose to UDP-galactose (also an LPS outer core component)[41]. Overexpression of *galU,* but not *galE*, *galF*, or *wecB,* completely reversed the *gnarl3* T5-sensitization phenotype in ECOR1, ECOR3, and ECOR22 (Figure 5f). suggesting

300 a specific interaction between *galU* and *gnarl3*. The same effect was observed with the
301 unrelated phage T4 (Supplementary Figure 5), indicating the *gnarl3-galU* interaction was phage-
302 agnostic. Overexpressing *galU* without *gnarl3* did not affect phage sensitivity except in
303 ECOR22, where it partially potentiated infection for unknown reasons. Coupled with LPS
304 electrophoresis and AP-MS data, this observation suggests a model where *gnarl3* inhibits GalU,
305 reducing the availability of UDP-glucose and resulting in a modification of the cell envelope that
306 potentiates phage adsorption. More broadly, it is notable that the most prevalent and potent
307 "counter-defense" phenotype for multiple, distinct AGs appears to hinge on removing barrier
308 defenses through diverse mechanisms.
309
310 **Barrier defenses and restriction systems provide layered immunity**
311 The existence of broad-spectrum counter-defense AGs underscores the importance of
312 barrier defenses. In strain ECOR21, transposon insertions within the O121-specific O-antigen
313 pathway[42] indeed sensitized the strain to multiple phages, but not T2 (Figure 6a). We wondered
314 if an additional defense mechanism in ECOR21 complemented barrier defenses by restricting
315 T2 infection (Figure 6b). In the AG screen, internal-protein IpII (*orf143*) partially sensitized
316 ECOR21 to T2 (Figure 5a), suggesting that T2 is blocked by an IpII-inhibited system in addition
317 to O121 O-antigen. Interestingly, although phage T2 itself does not encode IpII, phage T4
318 encodes IpII as well as the only internal-protein from T-even phages with a known function, IpI.
319 IpI is packaged into the phage capsid and disables canonical Type IV R-M (consisting of two
320 enzymes, GmrS and GmrD) upon co-injection with phage DNA[43].
321 To identify the IpII-inhibited system, we eliminated ECOR21 O-antigen by deleting
322 *wecA*[44] (Figure 6c) and repeated the transposon screen, looking for mutants in the
323 ECOR21Δ*wecA* background that now became sensitive to T2. We obtained several transposon
324 insertions in an atypical BREX system (Figure 6d) in which a fused Type IV R-M enzyme
325 GmrSD (also known as BrxU) is embedded[45]. Type IV R-M systems target glucosyl-hmC
326 modified DNA of T-even phages (T2, T4, T6)[43]. The entire ECOR21 BREX-GmrSD system
327 cloned into a lab strain and expressed from its native promoter blocked T2, but an active-site
328 mutant in GmrSD (GmrSD*) did not (Figure 6e). Inactivation of the BREX system (Δ*pglX*) had
329 no impact on T2, confirming that GmrSD causes T2 inhibition and is the likely target of IpII.
330 Consistent with this conclusion, expression of IpII, but neither Ocr (anti-Type I R-M and anti-
331 BREX[29]) nor IpI (which inhibits canonical GmrS-GmrD two component Type IV R-M, but not the
332 fused GmrSD[37,43]), abolished GmrSD defense (Figure 6e).
333 To test sufficiency, we expressed GmrSD or the catalytic mutant GmrSD* alone.
334 Overexpressing GmrSD was sufficient to block all T-even phages, while the catalytic mutant had
335 no effect (Figure 6f). Leaky expression of GmrSD from the uninduced plasmid still inhibited
336 replication of T2 and T6 (which do not encode IpII), but not the IpII-encoding phage T4. When
337 IpII was additionally expressed, T2 and T6 plaquing was restored, whereas plaquing of T4 did
338 not improve further (Figure 6g). Of the many described T-even phage internal proteins[12], IpII is
339 only the second (after IpI) with an identified target. The observed redundancy between barrier
340 defenses and restriction systems likely obscures phenotypes of other anti-restriction AGs as
341 well.
342
343 **Conclusion**
344
345 Our strategy for identifying AGs and their interactions with host defenses can tease apart
346 the layers of bacterial anti-phage immunity, providing a holistic view of the most salient
347 defenses across multiple strains. Deploying this platform in pathogenic bacteria will help tailor
348 phage therapy approaches to overcome species- and strain-specific antiviral mechanisms. In *E.*
349 *coli*, we discovered a diverse array of AG functions that interact with multilayered host immunity

at every level. AG-encoded proteins modify the cell surface to block viral entry, disable antiviral restriction systems inside the cell, and trigger PCD mechanisms in the host and its prophages, while cellular decoy immunity guards the host against phage counter-defense strategies. We show here how restriction systems work in tandem with barrier defenses on the cell surface, and with Abi/decoy guardians within the cell to protect the host.

We report surprising exaptation of well-studied R-M systems that are triggered by anti-restriction proteins to induce PCD. Our results further expand the conceptual framework of layered anti-phage defense in bacteria, whereby anti-anti-restriction mechanisms such as retrons[23], PARIS[24], and PrrC[25] provide backup immunity via Abi. In particular, EcoR22I combines both R-M and PCD functions, and Ronin is a PCD mechanism that mimics an R-M system to thwart phage anti-R-M AGs under the decoy immunity strategy previously identified only in eukaryotes. Thus, experimentally assaying AGs in wild bacterial strains can uncover defense functions that cannot be identified by comparison with known defense or counter-defense factors. For instance, EcoR22I does not have any domains currently known to be involved in PCD. Similarly, *ronA*-like toxins have no previously characterized links to R-M.

Assaying AGs individually in diverse hosts also directly identifies otherwise elusive triggers of defense systems. Anti-phage defense systems have been shown to detect highly conserved features of phages (nucleic acids, capsids, terminases, or portal proteins)[15-18]. We show that a variety of poorly conserved phage AGs can also trigger defense mechanisms in the host and its prophages. Non-essential gene products might seem ill-suited for this role because the virus could simply lose the gene to escape detection. However, all AGs we found to trigger PCD are associated with common counter-defense functions, trapping the virus into a Scylla vs Charybdis catch where a counter-defense AG triggers PCD upon infection, but losing that AG renders the virus vulnerable to immunity. The strategies viruses employ to navigate this treacherous narrows are currently unknown and might be an important force in virus evolution. The host counterpoints to virus AGs were likely overlooked because many viruses encoding inhibitors and triggers of specialized host defenses can be difficult to procure and culture. Advanced technologies for DNA synthesis combined with rapidly growing metagenomic sequence databases are starting to paint a more complete picture of virus-host coevolution.

379 **References**
380

381 1 Cumby, N., Davidson, A. R. & Maxwell, K. L. The moron comes of age. *Bacteriophage*
382 **2**, 225-228 (2012). https://doi.org:10.4161/bact.23146

383 2 Roucourt, B. & Lavigne, R. The role of interactions between phage and bacterial proteins
384 within the infected cell: a diverse and puzzling interactome. *Environmental microbiology*
385 **11**, 2789-2805 (2009). https://doi.org:10.1111/j.1462-2920.2009.02029.x

386 3 Koonin, E. V. & Wolf, Y. I. Evolution of microbes and viruses: a paradigm shift in
387 evolutionary biology? *Front Cell Infect Microbiol* **2**, 119 (2012).
388 https://doi.org:10.3389/fcimb.2012.00119

389 4 Koonin, E. V., Makarova, K. S. & Wolf, Y. I. Evolutionary Genomics of Defense
390 Systems in Archaea and Bacteria. *Annual review of microbiology* **71**, 233-261 (2017).
391 https://doi.org:10.1146/annurev-micro-090816-093830

392 5 Kruger, D. H., Schroeder, C., Hansen, S. & Rosenthal, H. A. Active protection by
393 bacteriophages T3 and T7 against E. coli B- and K-specific restriction of their DNA. *Mol*
394 *Gen Genet* **153**, 99-106 (1977). https://doi.org:10.1007/BF01036001

395 6 Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes
396 that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429-432 (2013).
397 https://doi.org:10.1038/nature11723

398 7 Kristensen, D. M., Mushegian, A. R., Dolja, V. V. & Koonin, E. V. New dimensions of
399 the virus world discovered through metagenomics. *Trends Microbiol* **18**, 11-19 (2010).
400 https://doi.org:10.1016/j.tim.2009.11.003

401 8 Juhala, R. J. *et al.* Genomic sequences of bacteriophages HK97 and HK022: pervasive
402 genetic mosaicism in the lambdoid bacteriophages. *Journal of molecular biology* **299**, 27-
403 51 (2000). https://doi.org:10.1006/jmbi.2000.3729

404 9 Comeau, A. M., Bertrand, C., Letarov, A., Tetart, F. & Krisch, H. M. Modular
405 architecture of the T4 phage superfamily: a conserved core genome and a plastic
406 periphery. *Virology* **362**, 384-396 (2007). https://doi.org:10.1016/j.virol.2006.12.031

407 10 Bondy-Denomy, J. *et al.* Prophages mediate defense against phage infection through
408 diverse mechanisms. *The ISME journal* **10**, 2854-2866 (2016).
409 https://doi.org:10.1038/ismej.2016.79

410 11 Tsao, Y. F. *et al.* Phage Morons Play an Important Role in Pseudomonas aeruginosa
411 Phenotypes. *Journal of bacteriology* **200** (2018). https://doi.org:10.1128/JB.00189-18

412 12 Repoila, F., Tetart, F., Bouet, J. Y. & Krisch, H. M. Genomic polymorphism in the T-
413 even bacteriophages. *The EMBO journal* **13**, 4181-4192 (1994).

414 13 Moffett, P. Using Decoys to Detect Pathogens: An Integrated Approach. *Trends Plant Sci*
415 **21**, 369-370 (2016). https://doi.org:10.1016/j.tplants.2016.04.002

416 14 Ochman, H. & Selander, R. K. Standard reference strains of Escherichia coli from natural
417 populations. *Journal of bacteriology* **157**, 690-693 (1984).
418 https://doi.org:10.1128/JB.157.2.690-693.1984

419 15 Lopatina, A., Tal, N. & Sorek, R. Abortive Infection: Bacterial Suicide as an Antiviral
420 Immune Strategy. *Annu Rev Virol* **7**, 371-384 (2020). https://doi.org:10.1146/annurev-
421 virology-011620-040628

422 16 Gao, L. A. *et al.* Prokaryotic innate immunity through pattern recognition of conserved
423 viral proteins. *Science* **377**, eabm4096 (2022). https://doi.org:10.1126/science.abm4096

424 17 Zhang, T. *et al.* Direct activation of a bacterial innate immune system by a viral capsid
425 protein. *Nature* **612**, 132-140 (2022). https://doi.org:10.1038/s41586-022-05444-z

426 18 Stokar-Avihail, A. *et al.* Discovery of phage determinants that confer sensitivity to
427 bacterial immune systems. *Cell* (2023). https://doi.org:10.1016/j.cell.2023.02.029

428 19 Roberts, G. A. *et al.* Exploring the DNA mimicry of the Ocr protein of phage T7. *Nucleic*
429 *acids research* **40**, 8129-8143 (2012). https://doi.org:10.1093/nar/gks516

430 20 McMahon, S. A. *et al.* Extensive DNA mimicry by the ArdA anti-restriction protein and
431 its role in the spread of antibiotic resistance. *Nucleic acids research* **37**, 4887-4897
432 (2009). https://doi.org:10.1093/nar/gkp478

433 21 Kruger, D. H., Reuter, M., Hansen, S. & Schroeder, C. Influence of phage T3 and T7
434 gene functions on a type III(EcoP1) DNA restriction-modification system in vivo. *Mol*
435 *Gen Genet* **185**, 457-461 (1982). https://doi.org:10.1007/BF00334140

436 22 Vasu, K. & Nagaraja, V. Diverse functions of restriction-modification systems in
437 addition to cellular defense. *Microbiol Mol Biol Rev* **77**, 53-72 (2013).
438 https://doi.org:10.1128/MMBR.00044-12

439 23 Millman, A. *et al.* Bacterial Retrons Function In Anti-Phage Defense. *Cell* **183**, 1551-
440 1561 e1512 (2020). https://doi.org:10.1016/j.cell.2020.09.065

441 24 Rousset, F. *et al.* Phages and their satellites encode hotspots of antiviral systems. *Cell*
442 *host & microbe* **30**, 740-753 e745 (2022). https://doi.org:10.1016/j.chom.2022.02.018

443 25 Penner, M., Morad, I., Snyder, L. & Kaufmann, G. Phage T4-coded Stp: double-edged
444 effector of coupled DNA and tRNA-restriction systems. *Journal of molecular biology*
445 **249**, 857-868 (1995). https://doi.org:10.1006/jmbi.1995.0343

446 26 Ofir, G. *et al.* DISARM is a widespread bacterial defence system with broad anti-phage
447 activities. *Nat Microbiol* **3**, 90-98 (2018). https://doi.org:10.1038/s41564-017-0051-0

448 27 Koonin, E. V. & Zhang, F. Coupling immunity and programmed cell suicide in
449 prokaryotes: Life-or-death choices. *Bioessays* **39**, 1-9 (2017).
450 https://doi.org:10.1002/bies.201600186

451 28 Murray, N. E. Type I restriction systems: sophisticated molecular machines (a legacy of
452 Bertani and Weigle). *Microbiol Mol Biol Rev* **64**, 412-434 (2000).
453 https://doi.org:10.1128/MMBR.64.2.412-434.2000

454 29 Isaev, A. *et al.* Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence.
455 *Nucleic acids research* **48**, 5397-5406 (2020). https://doi.org:10.1093/nar/gkaa290

456 30 Murphy, K. C. Bacteriophage P22 Abc2 protein binds to RecC increases the 5' strand
457 nicking activity of RecBCD and together with lambda bet, promotes Chi-independent
458 recombination. *Journal of molecular biology* **296**, 385-401 (2000).
459 https://doi.org:10.1006/jmbi.1999.3486

460 31 Ghisotti, D. *et al.* Nonessential region of bacteriophage P4: DNA sequence, transcription,
461 gene products, and functions. *Journal of virology* **64**, 24-36 (1990).
462 https://doi.org:10.1128/JVI.64.1.24-36.1990

463 32 Kulikov, E. E. *et al.* Equine Intestinal O-Seroconverting Temperate Coliphage Hf4s:
464 Genomic and Biological Characterization. *Applied and environmental microbiology* **87**,
465 e0112421 (2021). https://doi.org:10.1128/AEM.01124-21

466 33 Allison, G. E. & Verma, N. K. Serotype-converting bacteriophages and O-antigen
467 modification in Shigella flexneri. *Trends Microbiol* **8**, 17-23 (2000).
468 https://doi.org:10.1016/s0966-842x(99)01646-7

469 34 Vander Byl, C. & Kropinski, A. M. Sequence of the genome of Salmonella bacteriophage
470 P22. *Journal of bacteriology* **182**, 6472-6481 (2000).
471 https://doi.org:10.1128/JB.182.22.6472-6481.2000

472 35 Newton, G. J. *et al.* Three-component-mediated serotype conversion in Pseudomonas
473 aeruginosa by bacteriophage D3. *Mol Microbiol* **39**, 1237-1247 (2001).
474 https://doi.org:10.1111/j.1365-2958.2001.02311.x

475 36 Losick, R. Isolation of a trypsin-sensitive inhibitor of O-antigen synthesis involved in
476 lysogenic conversion by bacteriophage epsilon-15. *Journal of molecular biology* **42**, 237-
477 246 (1969). https://doi.org:10.1016/0022-2836(69)90040-0

478 37 Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature*
479 *reviews. Microbiology* **8**, 317-327 (2010). https://doi.org:10.1038/nrmicro2315

480 38 Weissborn, A. C., Liu, Q., Rumley, M. K. & Kennedy, E. P. UTP: alpha-D-glucose-1-
481 phosphate uridylyltransferase of Escherichia coli: isolation and DNA sequence of the
482 galU gene and purification of the enzyme. *Journal of bacteriology* **176**, 2611-2618
483 (1994). https://doi.org:10.1128/jb.176.9.2611-2618.1994

484 39 Genevaux, P., Bauda, P., DuBow, M. S. & Oudega, B. Identification of Tn10 insertions
485 in the rfaG, rfaP, and galU genes involved in lipopolysaccharide core biosynthesis that
486 affect Escherichia coli adhesion. *Arch Microbiol* **172**, 1-8 (1999).
487 https://doi.org:10.1007/s002030050732

488 40 Marolda, C. L. & Valvano, M. A. The GalF protein of Escherichia coli is not a UDP-
489 glucose pyrophosphorylase but interacts with the GalU protein possibly to regulate
490 cellular levels of UDP-glucose. *Mol Microbiol* **22**, 827-840 (1996).
491 https://doi.org:10.1046/j.1365-2958.1996.01531.x

492 41 Pierson, D. E. & Carlson, S. Identification of the galE gene and a galE homolog and
493 characterization of their roles in the biosynthesis of lipopolysaccharide in a serotype O:8
494 strain of Yersinia enterocolitica. *Journal of bacteriology* **178**, 5916-5924 (1996).
495 https://doi.org:10.1128/jb.178.20.5916-5924.1996

496 42 Fratamico, P. M., Briggs, C. E., Needle, D., Chen, C. Y. & DebRoy, C. Sequence of the
497 Escherichia coli O121 O-antigen gene cluster and detection of enterohemorrhagic E. coli
498 O121 by PCR amplification of the wzx and wzy genes. *J Clin Microbiol* **41**, 3379-3383
499 (2003). https://doi.org:10.1128/JCM.41.7.3379-3383.2003

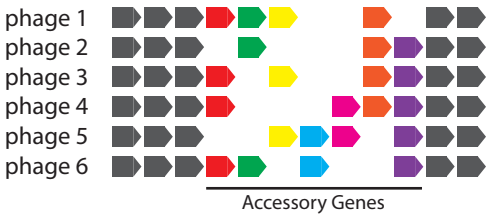
500 43 Rifat, D., Wright, N. T., Varney, K. M., Weber, D. J. & Black, L. W. Restriction
501 endonuclease inhibitor IPI* of bacteriophage T4: a novel structure for a dedicated target.
502 *Journal of molecular biology* **375**, 720-734 (2008).
503 https://doi.org:10.1016/j.jmb.2007.10.064

504 44 Lehrer, J., Vigeant, K. A., Tatar, L. D. & Valvano, M. A. Functional characterization and
505 membrane topology of Escherichia coli WecA, a sugar-phosphate transferase initiating
506 the biosynthesis of enterobacterial common antigen and O-antigen lipopolysaccharide.
507 *Journal of bacteriology* **189**, 2618-2628 (2007). https://doi.org:10.1128/JB.01905-06

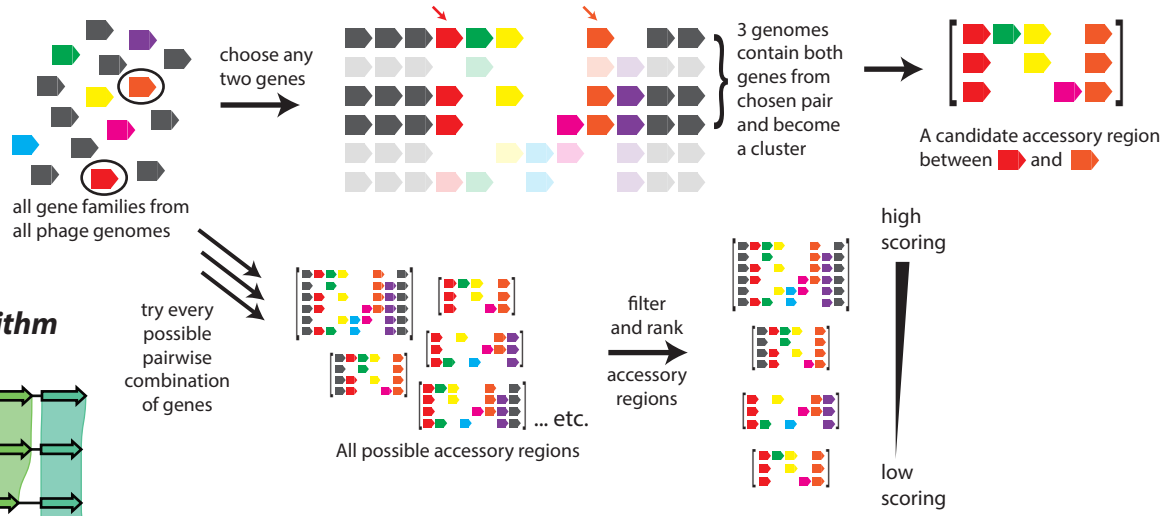
508 45 Picton, D. M. *et al.* The phage defence island of a multidrug resistant plasmid uses both
509 BREX and type IV restriction for complementary protection from viruses. *Nucleic acids*
510 *research* **49**, 11257-11273 (2021). https://doi.org:10.1093/nar/gkab906

511

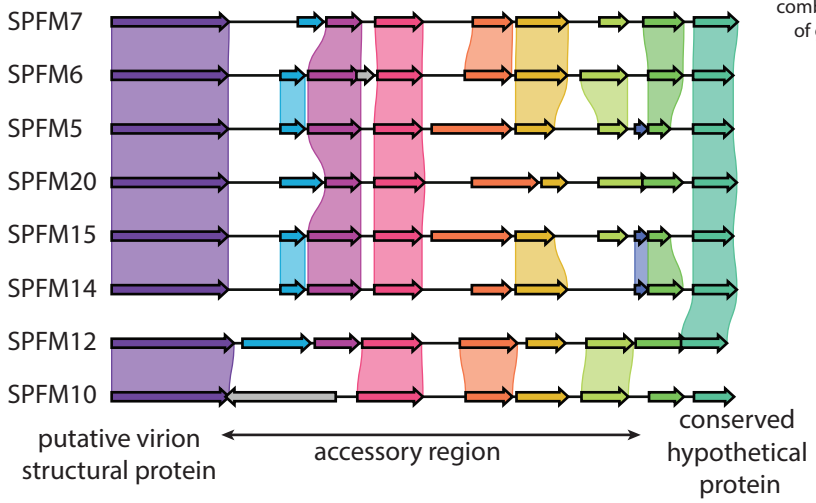
A. Core vs. Accessory genes in a hypothetical phage cluster



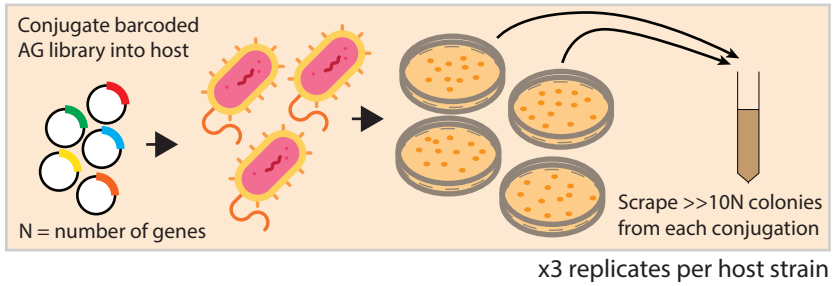
B. Algorithm to find accessory regions without a priori genome clustering

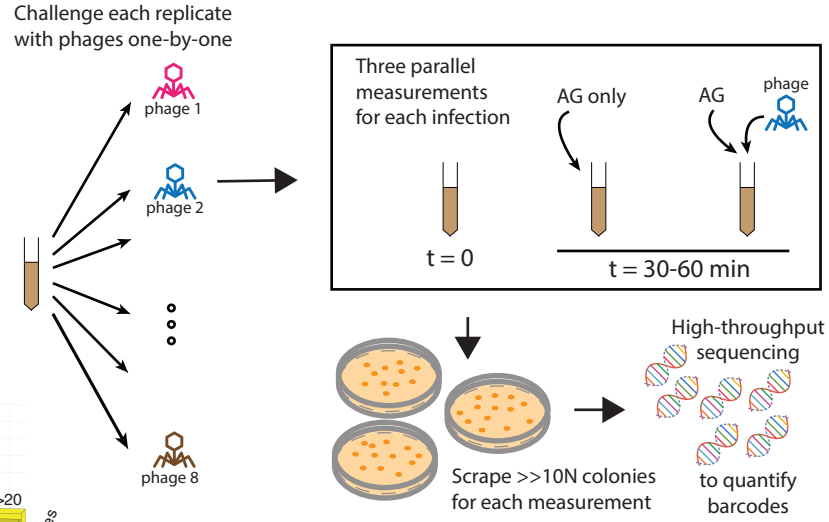


C. Sample accessory region found by the algorithm

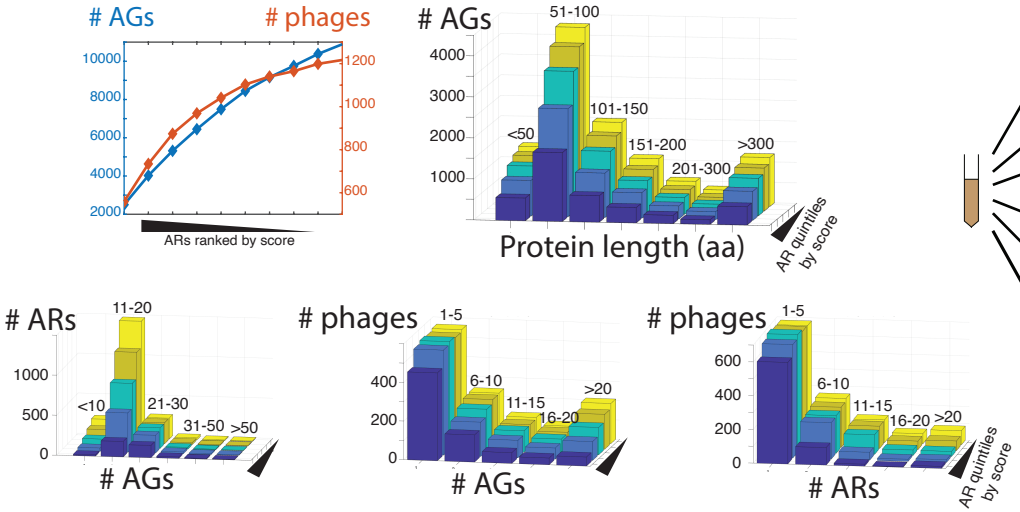


E. Screen to assay AG phenotypes in parallel

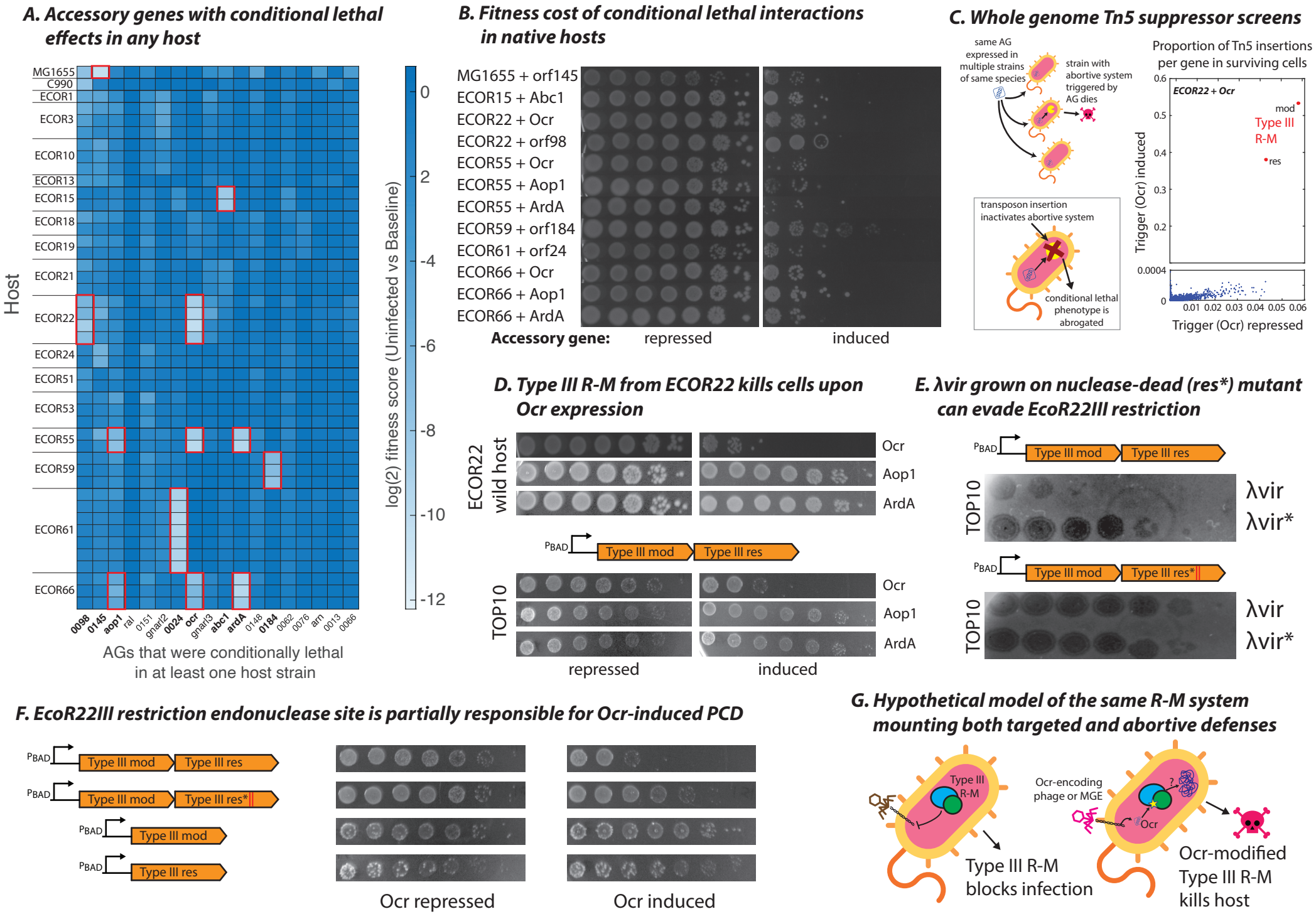




D. Characteristics of ARs and AGs

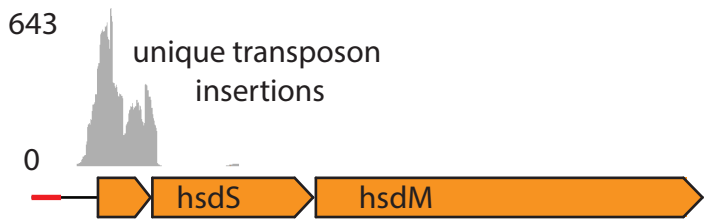


513 **Figure 1. Platform for identifying and testing phage AGs. (A)** Schematic of a genomic region
514 showing core (shown in grey) and accessory (in various colors) genes in a hypothetical phage
515 genome cluster. **(B)** Schematic of algorithm to exhaustively identify accessory regions using
516 pairwise combinations of genes, without *a priori* genome clustering. **(C)** A sample high-scoring
517 AR in a family of *Salmonella* phages. Purple shading denotes >80% and green denotes >90%
518 nucleotide identity. **(D)** Statistical features of ARs and AGs. 2D chart depicts saturation curves
519 (cumulative distribution functions) of unique AGs (left axis) and phage genomes (right axis)
520 represented in ~2000 non-redundant ARs ranked by score on the X-axis. 3D bar graphs show
521 various distributions (length of AGs, number of AGs per AR, number of AGs per phage, and
522 number of ARs per phage). Color-coded histograms in each bar graph are cumulative i.e., the
523 first histogram (dark blue) shows the top 20% highest-scoring ARs, the next (light blue) represents
524 top 40% and so on. **(E)** Schematic of the AG screen.
525

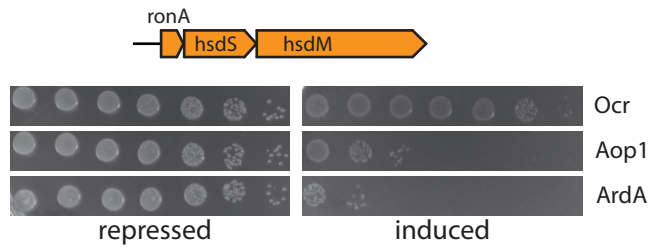

A. Accessory genes with conditional lethal effects in any host
Host
MG1655
C990
ECOR1
ECOR3
ECOR10
ECOR13
ECOR15
ECOR18
ECOR19
ECOR21
ECOR22
ECOR24
ECOR51
ECOR53
ECOR55
ECOR59
ECOR61
ECOR66
0098 0145 aop1 ral 0151 gnarl2 0024 ocr gnarl3 abc1 ardA 0148 0184 0062 0076 arn 0013 0066
AGs that were conditionally lethal in at least one host strain
log(2) fitness score (Uninfected vs Baseline)
0 -2 -4 -6 -8 -10 -12
B. Fitness cost of conditional lethal interactions in native hosts
MG1655 + orf145
ECOR15 + Abc1
ECOR22 + Ocr
ECOR22 + orf98
ECOR55 + Ocr
ECOR55 + Aop1
ECOR55 + ArdA
ECOR59 + orf184
ECOR61 + orf24
ECOR66 + Ocr
ECOR66 + Aop1
ECOR66 + ArdA
Accessory gene: repressed induced
C. Whole genome Tn5 suppressor screens
same AG expressed in multiple strains of same species
strain with abortive system triggered by AG dies
transposon insertion inactivates abortive system
conditional lethal phenotype is abrogated
Proportion of Tn5 insertions per gene in surviving cells
ECOR22 + Ocr
mod
Type III R-M
res
Trigger (Ocr) induced
Trigger (Ocr) repressed
D. Type III R-M from ECOR22 kills cells upon Ocr expression
ECOR22 wild host
TOP10
PBAD
Type III mod
Type III res
Ocr
Aop1
ArdA
repressed induced
E. λvir grown on nuclease-dead (res*) mutant can evade EcoR22III restriction
Type III res*
TOP10
λvir
λvir*
F. EcoR22III restriction endonuclease site is partially responsible for Ocr-induced PCD
Ocr repressed
Ocr induced
G. Hypothetical model of the same R-M system mounting both targeted and abortive defenses
Type III R-M
Ocr-encoding phage or MGE
Ocr
Type III R-M blocks infection
Ocr-modified Type III R-M kills host

527 **Figure 2. AGs that trigger programmed cell death. (A)** Heatmap of log2-transformed fitness
528 scores comparing induced and uninduced samples without phage infection. Of 200 AGs tested,
529 only AGs that exhibited lethality in some (but not all) hosts are shown. Red boxes/bold names
530 highlight conditional-lethal AGs selected for further study. Named AGs are *orf7:aop1, orf87:ral,*
531 *orf63:gnarl2, orf1:ocr, orf92:gnarl3, orf116:abc1, orf169:ardA* and *orf2:arn*. **(B)** Host-AG
532 combinations in red boxes from (C) tested individually for AG-induced lethality (serial tenfold
533 dilutions of saturated culture; AGs repressed: left, induced: right). **(C)** Schematic depicts
534 hypothetical strain-specific Abi system being triggered by AGs, and inactivated in follow-up
535 transposon knockout screens (inset). Data from a representative screen with a Tn5 library
536 constructed from ECOR22 with chromosomally encoded *ocr*. Proportion of transposon insertions
537 per gene in surviving cells with *ocr* repressed is plotted on the X-axis, and with *ocr* induced on
538 the Y-axis. Y-axis interrupted to resolve genes other than Type III R-M *mod*, *res*. **(D)** Abortive
539 phenotype of EcoR22I in its native host, and reconstituted when cloned onto a plasmid and
540 expressed in TOP10 (AGs repressed: left, induced: right)**. (E)** Plaque assay showing R-M
541 phenotype of EcoR22I against methylated and unmethylated λvir in TOP10. λvir* were methylated
542 by passaging in a TOP10 strain with nuclease-dead EcoR22I (res* double mutant; D1024A,
543 D1043A) and tested for their ability to evade restriction by wildtype or res* EcoR22I. **(F)** As in (D),
544 PCD activity of EcoR22I variants in TOP10 (wildtype, nuclease dead (res*), and *mod*, *res* genes
545 expressed individually). **(G)** Hypothetical model of EcoR22I R-M and PCD activities. EcoR22I kills
546 its host in the presence of Ocr, partially relying on its nuclease activity. "pBAD" indicates the
547 construct was overexpressed from the plasmid in (F-H).
548

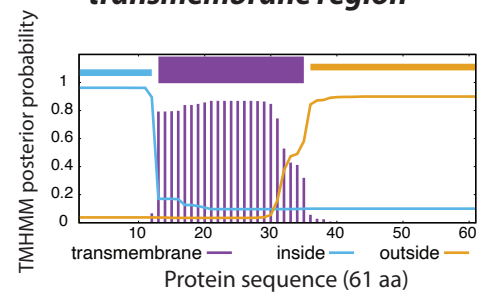
A. Transposon insertions that inactivate PCD in ECOR55 despite expression of anti-R-M genes



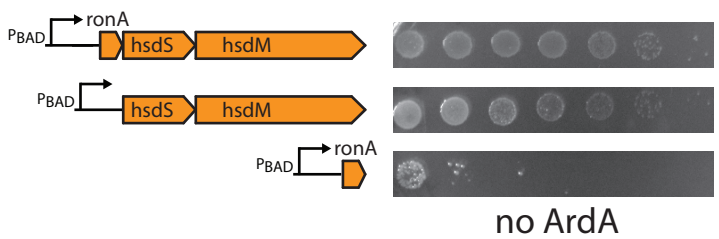
B. Ronin system from ECOR55 kills cells upon aop1 or ardA induction



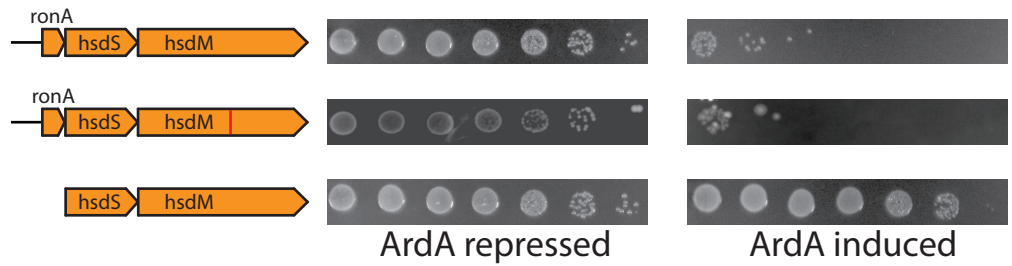
C. RonA contains a predicted transmembrane region



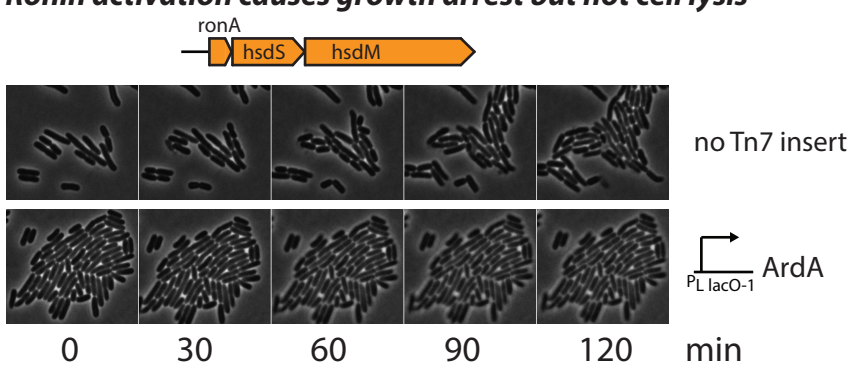
D. RonA is toxic upon overexpression but not with hsdMS present



E. RonA is necessary for ArdA-induced PCD activity



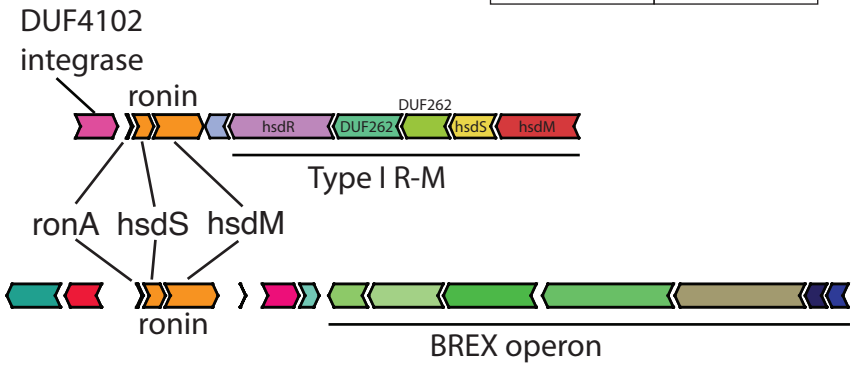
F. Ronin activation causes growth arrest but not cell lysis



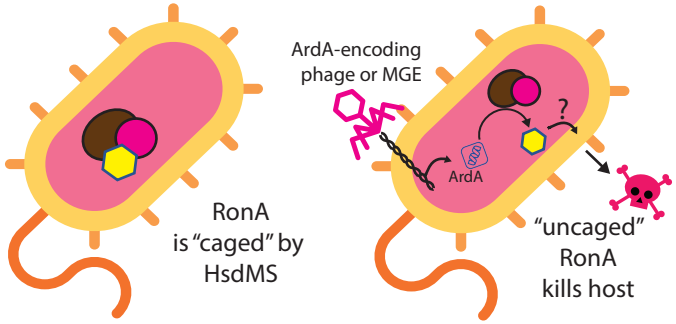
G. RonA-associated type I R-M systems (ronin) lack the restriction endonuclease

genomes with *ronA* (84) → *ronA* linked to *hsdMS* (84) → DUF4102 integrase upstream (58) / lacking full integrase upstream (26)

	present in *E. coli*	full Type I R-M anywhere
DUF4102 integrase upstream (58)	0	55
lacking full integrase upstream (26)	26	6

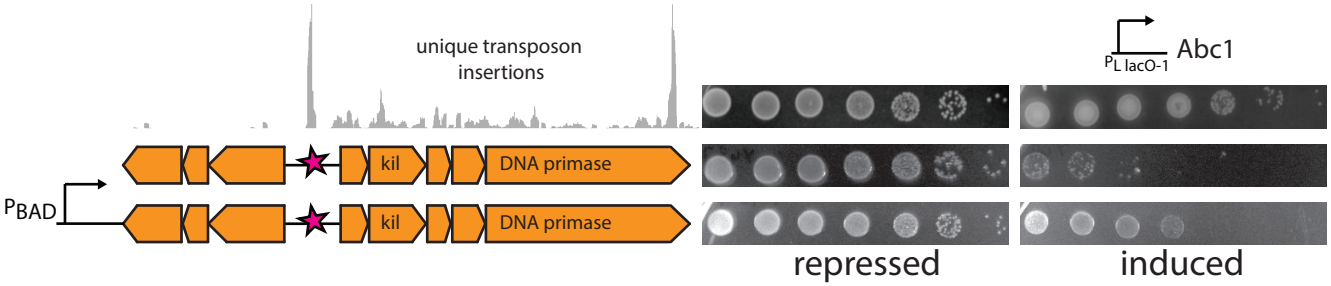


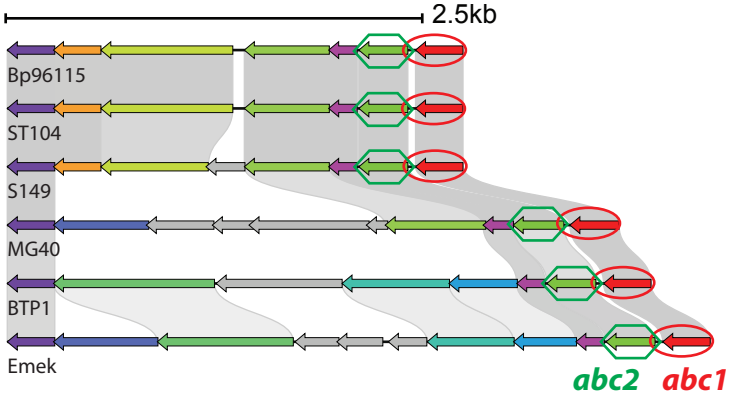
H. Hypothetical model of RonA "uncaging"



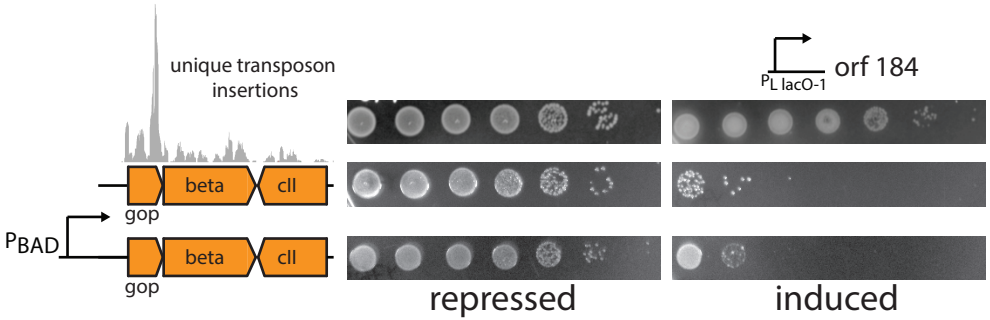
550 **Figure 3. Decoy immunity in bacteria. (A)** Distribution of unique transposon insertions
551 recovered from survivors in ECOR55 Tn5 libraries upon *aop1* induction, mapped to Ronin. **(B)**
552 Schematic of Ronin system from ECOR55. Abortive phenotype reconstituted in TOP10 upon
553 expression of Ronin from a plasmid (serial tenfold dilutions of saturated culture; AGs repressed:
554 left, induced: right). **(C)** Prediction of transmembrane character along the 61 aa RonA sequence.
555 **(D)** Lethality of Ronin components tested when overexpressed (indicated as "pBAD") in TOP10
556 without AGs. **(E)** As in (B), genetic requirements for the Abi phenotype of Ronin. Red bar in *hsdM*
557 (second row) denotes the F287G active-site mutation. **(F)** Time-lapse imaging of TOP10 cells
558 expressing Ronin with (bottom) or without (top) chromosomally-integrated ArdA. **(G)** Two
559 prevalent architectures of *ronA*-linked Type I R-M-like Abi system (ronin), based on the presence
560 of an upstream DUF4102 integrase. Number of (*E. coli* and other) genomes with a full Type I R-
561 M system (with *hsdR, hsdM, hsdS* genes) anywhere in the genome is shown for each Ronin
562 architecture. **(H)** Schematic depicting hypothetical RonA "uncaging" when ArdA is present.
563

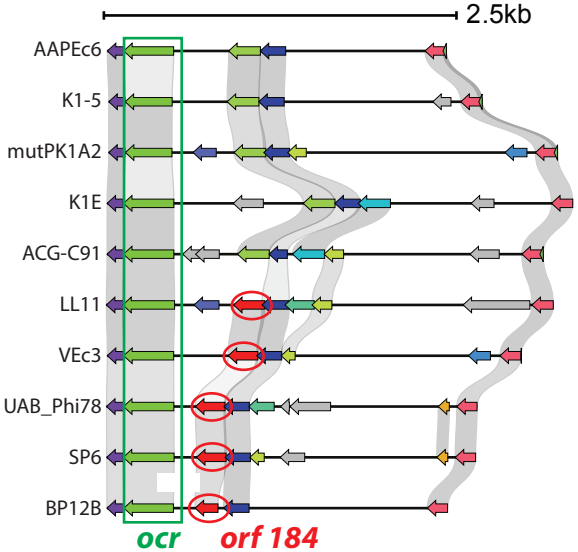
A. P4 satellite prophage from ECOR15 arrests TOP10 upon abc1 induction



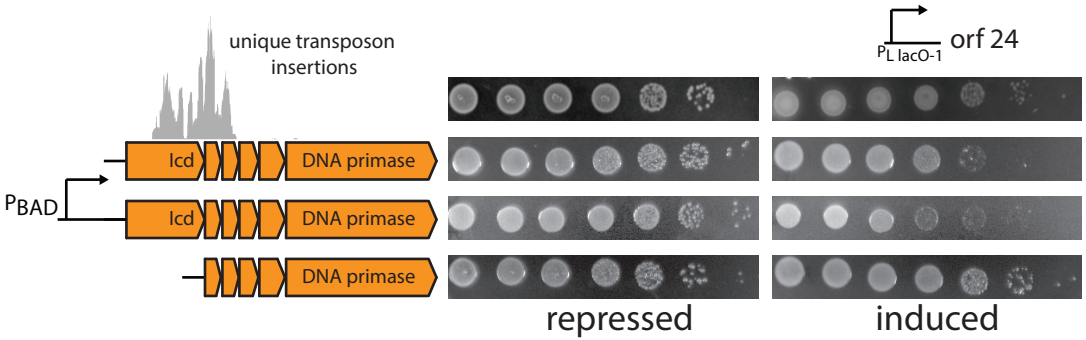


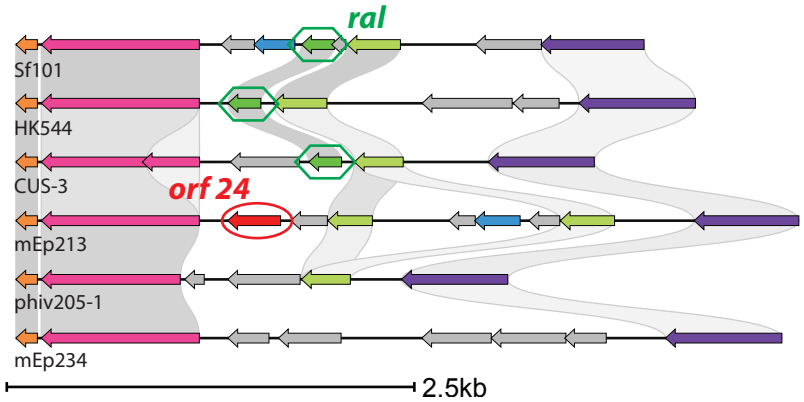
B. gop-beta-cII system from ECOR59 arrests TOP10 upon orf184 induction





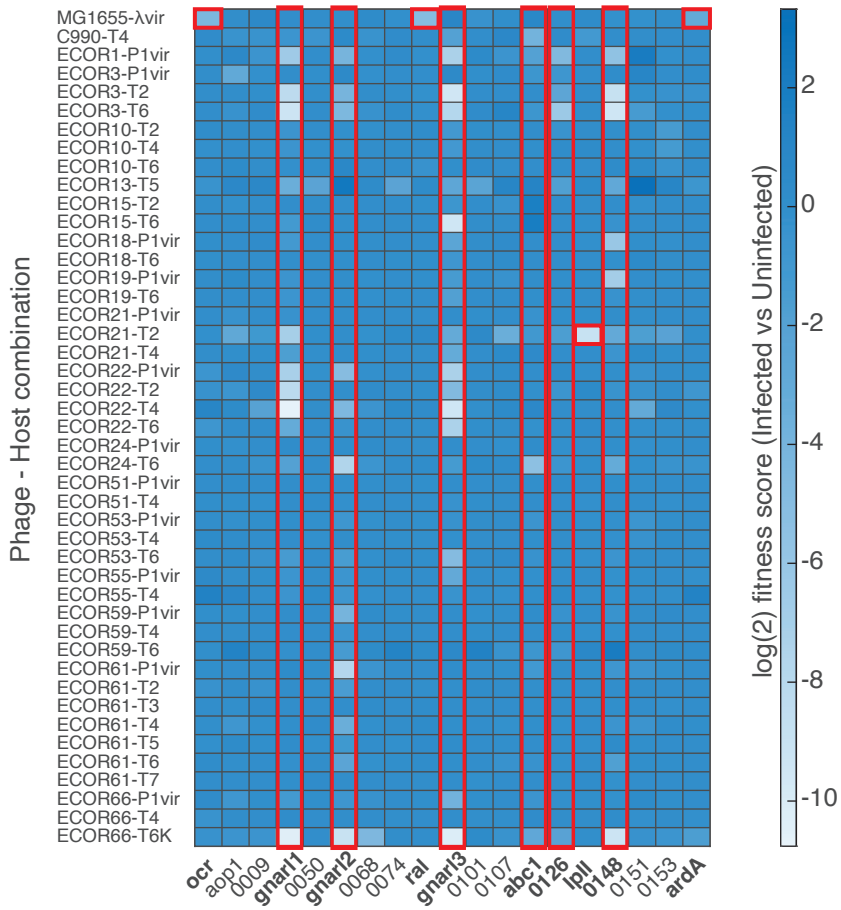
C. Novel prophage-encoded system from ECOR61 arrests TOP10 upon orf24 induction





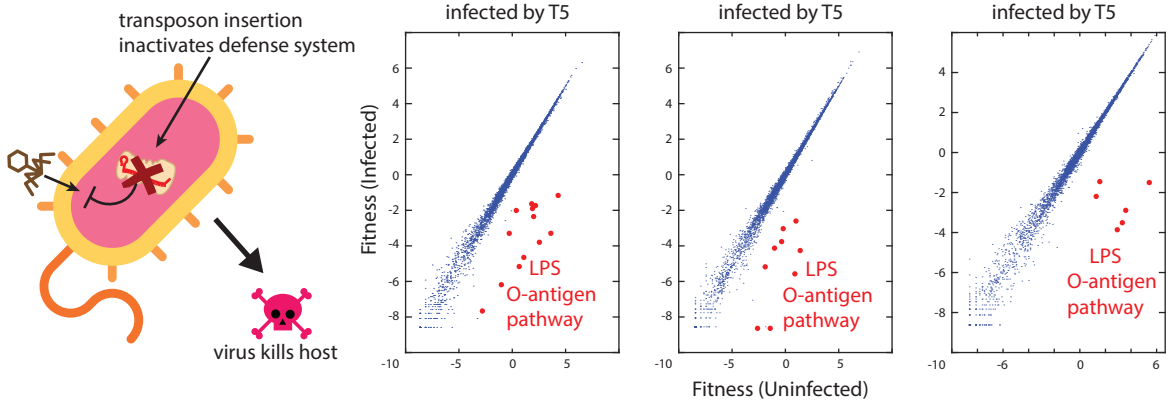
565 **Figure 4. AGs that trigger prophage-encoded PCD mechanisms. (A)** Abortive phenotype of
566 P4 satellite prophage from ECOR15 cloned onto a plasmid and expressed in TOP10. Abc1 is
567 expressed from a single-copy chromosomal insertion (serial tenfold dilutions of saturated culture;
568 AGs repressed: left, induced: right). Star marks an intergenic region that was strongly targeted in
569 transposon suppressor screens. **(B)** As in (A), PCD phenotype of gop-beta-cII system from
570 ECOR59 upon expression of *orf184*. **(C)** As in (A), PCD phenotype of novel prophage-encoded
571 superinfection exclusion system from ECOR61 upon expression of *orf24*. "pBAD" indicates where
572 the constructs were overexpressed in (A-C). Distribution of unique transposon insertions
573 recovered from survivors in Tn5 libraries upon trigger AG induction are mapped to each PCD
574 locus in (A-C). Accessory Regions where trigger AGs in (A-C) were found (*orf116*:*abc1*, *orf184*,
575 *orf24*; highlighted with red ovals), with co-occurring counter-defense AGs (anti-RecBCD *abc2*,
576 anti-R-M *ocr*, anti-R-M *ral*; indicated in green boxes) are shown in the respective panels.
577

A. Accessory genes with counter-defense phenotypes in any host

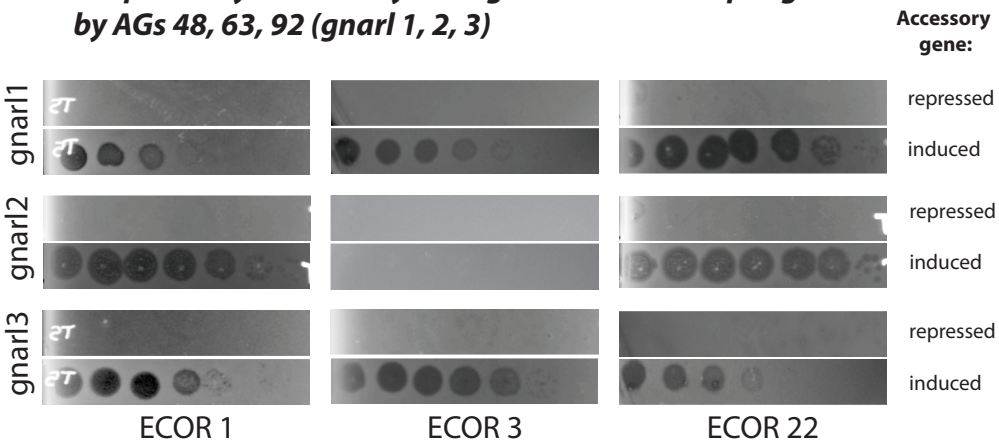


AGs that had a counter-defense effect in at least one host strain

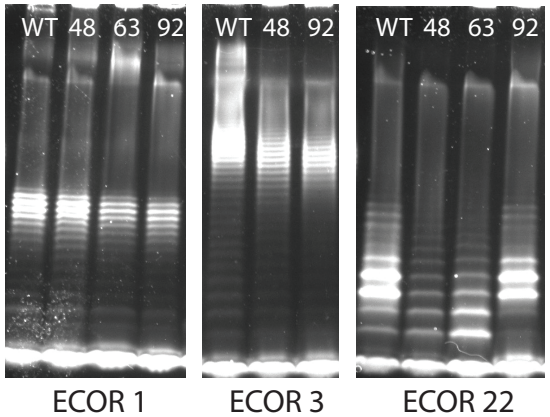
B. Whole genome Tn5-knockout screening to find phage exclusion mechanisms



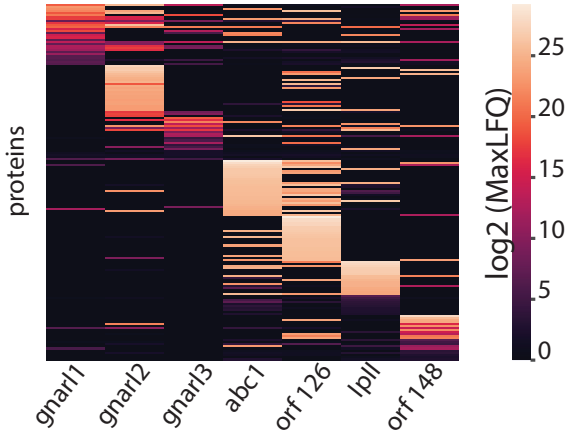
C. Plaque assays show very strong sensitization to phage T5 by AGs 48, 63, 92 (gnarl 1, 2, 3)



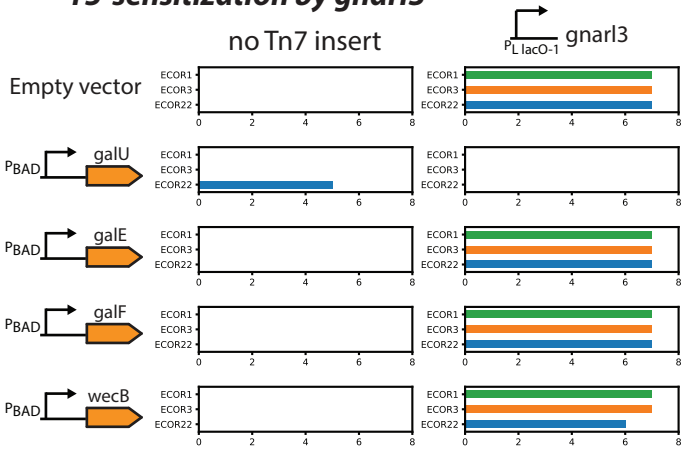
D. Gel electrophoresis of LPS from wild E. coli expressing AGs 48, 63, 92 (gnarl 1, 2, 3)



E. Heatmap of enriched host-proteins upon AP-MS of broad-spectrum phage AGs

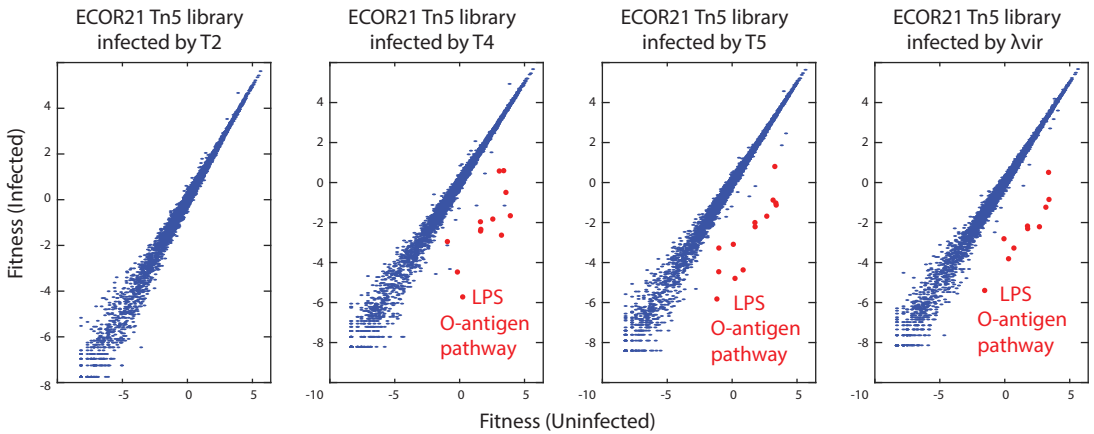


F. Over-expression of galU reverses T5-sensitization by gnarl3

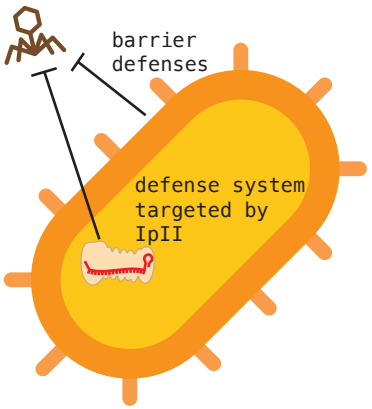


579 **Figure 5. AGs that sensitize hosts to infection by modifying barrier defenses. (A)**
580 Heatmap of log2-transformed fitness scores comparing infected and uninfected samples. Of 200
581 AGs tested, only AGs with a host-sensitizing phenotype are shown on the X-axis. Red
582 boxes/bold names highlight counter-defense AGs selected for further study. Named AGs are
583 *orf1:ocr, orf7:aop1, orf48:gnarl1, orf63:gnarl2, orf87:ral, orf92:gnarl3, orf116:abc1, orf143:lplI*
584 and *orf169:ardA*. **(B)** Schematic of follow-up transposon knockout screens, performed in the
585 native host without AGs. Graphs show fitness of transposon-mediated knockouts in infected (Y-
586 axis) and uninfected (X-axis) Tn5-libraries from three representative hosts challenged by T5.
587 Red dots are gene-disruptions that lower bacterial fitness upon infection (e.g., disruptions in O-
588 antigen biosynthesis). **(C)** Plaque assays with phage T5 and hosts from (D) with *gnarl* genes
589 repressed (top) or induced (bottom). **(D)** Lipopolysaccharide (LPS) visualized by electrophoresis
590 from host-AG combinations in (E) where AGs produced a phage-sensitizing effect. **(E)** Log2-
591 transformed protein representation scores (MaxLFQ) for the top 30 enriched host proteins upon
592 expression of AGs. **(F)** log10-transformed T5 infection scores in wild hosts ECOR1 (green),
593 ECOR3 (orange), ECOR22 (blue) with or without UDP-glucose biosynthesis pathway genes
594 *galU, galE, galF,* and ECA precursor *wecB* cloned onto a plasmid and overexpressed.
595

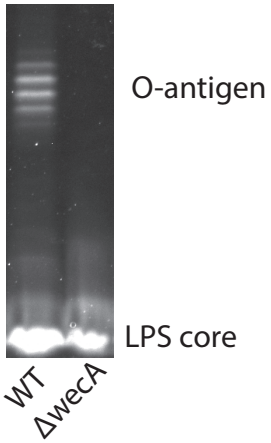
A. Whole genome Tn5-knockout screening in ECOR21 with various phages



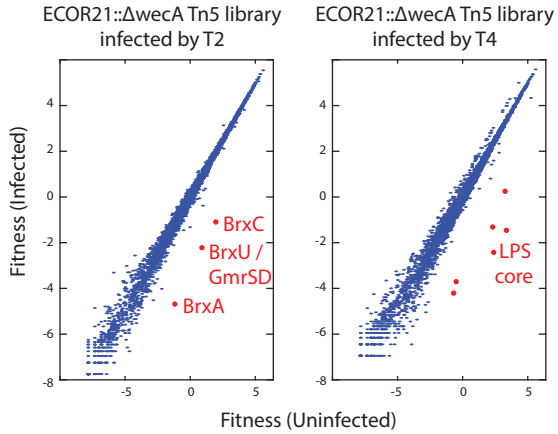
B. ECOR21 might block T2 at two levels simultaneously



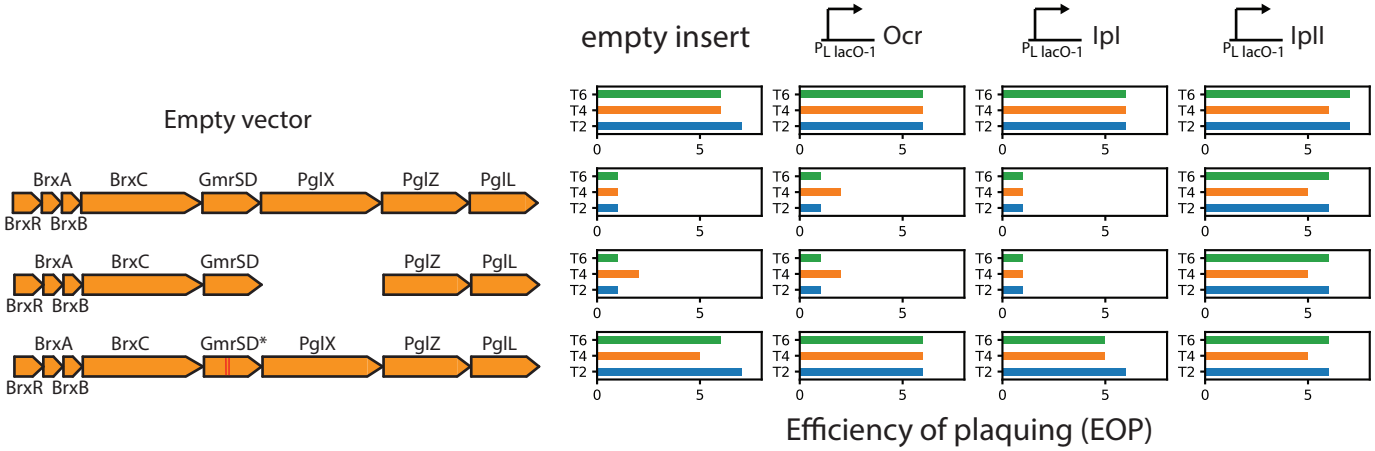
C. O-antigen removed in ECOR21::ΔwecA



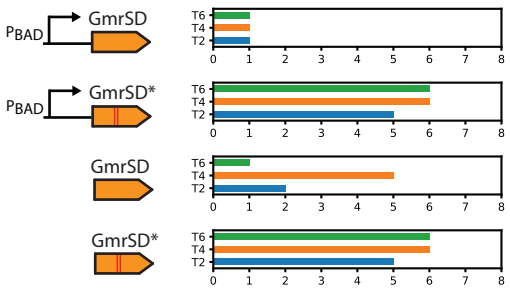
D. Tn5 screens after removing O-antigen layer



E. BREX-operon configurations tested against T-even phages with various inhibitors



F. GmrSD is sufficient to confer resistance against T-even phages



G. T4 naturally encodes IpII and does not require additional IpII to overcome basal levels of GmrSD restriction

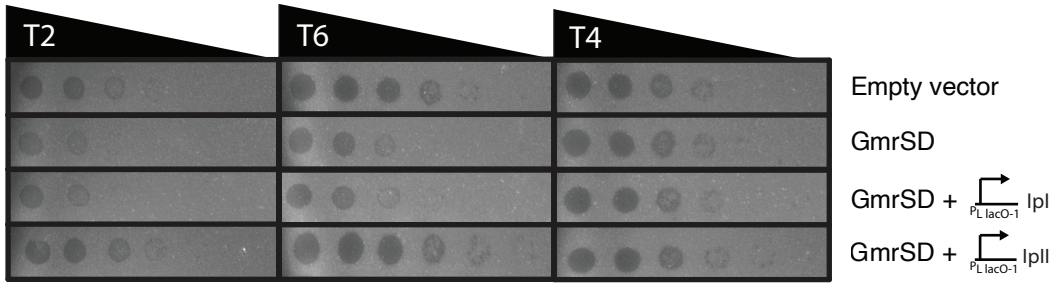


Figure 6. AGs that block canonical antiviral defenses. (A) Fitness of transposon-mediated knockouts in infected and uninfected Tn5-libraries from ECOR21 challenged by various phages. **(B)** Schematic of ECOR21 defense against T2 through both barrier defenses and a defense system which is putatively blocked by *orf143*/IpII. **(C)** LPS visualized from ECOR21 with the O-antigen biosynthesis initiator *wecA* deleted by allelic exchange. **(D)** Whole-genome transposon-knockout screens as in (A) but with the ECOR21::Δ*wecA*. **(E)** Variants of the BREX/GmrSD operon cloned onto a plasmid along with their native promoters and tested in a lab strain of *E. coli* (TOP10) against T-even phages (T2, T4, T6). Removal of the PglX gene inactivates BREX. Active-site mutations that disable GmrSD are shown as two parallel red bars (GmrSD* double mutant; D474A, H475A). Ocr, IpI and IpII are expressed from single-copy chromosomal insertions. Bar graphs show log10 infection scores. **(F)** GmrSD and its inactive variant GmrSD* cloned onto a plasmid and tested in TOP10 for restriction activity against T-even phages. "pBAD" indicates the defense construct was overexpressed from the plasmid. **(G)** Plaque assays with TOP10 cells with uninduced GmrSD (i.e., not overexpressed) challenged by T-even phages.

612 **Methods**
613
614 **Algorithm to identify phage accessory regions**
615 1706 non-redundant Enterobacteriophage genomes were obtained (with help from Dr.
616 Andrew Millard, Leicester) from public databases. These phages were both lytic and lysogenic,
617 but were sequenced and assembled from cultured isolates (i.e. prophages and metagenomic
618 samples were not included). All annotated protein records from all the phages were assigned
619 unique numeric IDs and clustered by MMseqs2[46]. ORF IDs in phage genomes were then
620 replaced by their MMseqs2 cluster representative IDs in each genome. Thus, each phage
621 genome can be represented as an ordered list of IDs of clustered proteins. The gene-
622 neighborhood of every ORF (up to 10 genes upstream and downstream) was extracted from
623 every virus. Given that phage genomic termini can be difficult to identify, every genome was
624 considered potentially circularly permuted. Allowing for circularization also ensures all gene-
625 neighborhoods were the same size and putative ARs towards ends of genomes would not be
626 missed. A 2D matrix of all possible pairwise combinations of genes was constructed, with every
627 matrix position containing lists of (a) the phages that each gene appears in, (b) the 10 upstream,
628 and (c) 10 downstream neighboring ORFs of each gene in every genome. We then traversed
629 the matrix and assessed every gene pair to identify all possible ARs, defined as the region
630 between a pair of genes that are both present in at least 6 different phage genomes but are no
631 more than 10 genes apart in any genome. Hence, every virus genome that contained both
632 genes of the gene pair was grouped into a possible virus cluster, and genes encoded between
633 the genes of the query pair were considered members of a possible AR. Regions trivially
634 containing the same genes in each participating genome were discarded. Genomes involved in
635 a putative AR were then compared pairwise at the nucleotide level using FastANI[47], and ARs
636 with divergent genomes were removed. In practice, this was done by pre-computing an all-vs-all
637 matrix of genome ANI and then traversing subsets of the matrix containing all the genomes
638 participating in the putative AR to ensure no cells were empty (FastANI produces no output
639 when very divergent genomes are compared). Various parameters were extracted from
640 surviving ARs. *N* is the number of unique genes in the AR. The vector *c* contains conservation
641 numbers (the number of different genomes that each gene appears in; also indicated on top of
642 each gene) for all the unique genes in this AR. The vector *l* contains the lengths (in numbers of
643 genes) of the AR in each genome. *CV* denotes coefficient of variation. *GCdev* is a vector
644 containing one minus the absolute deviation in GC content of each accessory gene from its
645 resident genome. A scaling factor is also included to normalize the score by the number of
646 phage genomes in the group that have at least one gene in the AR. ARs were scored according
647 to the formula below.

648 $$ARscore = \frac{N * median(GCdev) * CV(c)}{CV(l) * mean(c)} * \frac{\#non_empty}{\#genomes}$$

649
650 To filter out ARs that were subsets of other ARs, we successively removed redundant,
651 low scoring ARs with at least 90% of genes that had already been encountered in higher scoring
652 ARs. The top 200 ARs were examined manually and 1-15 AGs were selected from 62 ARs
653 (Supplementary Files 2,3). A worked example of the code is provided in Supplementary File 1.
654
655 **Strains and Vectors**
656 The *E. Coli* Reference (ECOR) collection of wild *E. coli* strains was obtained from the
657 STEC Center (Michigan State University). Lab strains were obtained as follows: MG1655 (from
658 Dr. Carol Gross, UCSF), C990 (from Dr. Ry Young, Texas A&M), DH5α (NEB), DH10ß/TOP10
659 (Life Technologies), BW25113 and BW25141 (from Dr. Vivek Mutalik, LBNL), and WM6026
660 (from Dr. Jason Peters, UW-Madison). All strains were cultured at 37°C with shaking (180-225

661 rpm) in lysogeny broth (LB) (10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl with 15 g/l agar for
662 plates) supplemented with 100 µg/ml of ampicillin/carbenicillin, 10-20 µg/ml gentamicin (except
663 WM6026, which yielded more consistent colony sizes with 50 µg/ml), 50 µg/ml kanamycin, and
664 0.5% w/v glucose, 0.1% w/v L-arabinose, and 1 mM IPTG (isopropyl-b-D-thiogalactopyranoside)
665 as needed. The WM6026 strain required diaminopimelic acid (DAP) for growth, which was
666 supplied at 300 µM. All strains were stored in 20% glycerol at -80°C for long term storage.
667 R6K-origin pTn7 plasmids (from Dr. Jason Peters, UW-Madison) – the
668 pJMP1039/pTn7C1 helper plasmid (carbenicillin) and the pJMP1360/pTn7C185 transposon
669 vector (gentamicin, carbenicillin) – were used for integration of phage AGs into the *E. coli*
670 genome. All host defense systems identified in the various screens were placed under the
671 araBAD promoter (arabinose inducible) in the pBAD Myc/His A vector (Invitrogen). Conjugative
672 allelic exchange was carried out using R6K-origin pKEK2201 (kanamycin) (from Karl E. Klose,
673 UT San Antonio). All plasmids were constructed by Gibson Assembly.
674 R6K-origin plasmids were maintained in BW25141 (routine cloning) and WM6026
675 (conjugative donor strain). All other plasmids were maintained in DH5α or TOP10. All plasmid
676 sequences were verified by Sanger or whole-plasmid sequencing (Plasmidsaurus, Primordium)
677 and are available upon request. Frequent plasmid dimerization during the Gibson Assembly
678 procedure (especially when using the NEBuilder reagent) was ameliorated by using 10-fold less
679 of one of the insert DNA fragments in a reaction involving at least 2 insert fragments (not
680 including the backbone).
681
682 **AG conjugation into host strains**
683 AGs were delivered to *E. coli* hosts by Tn7 transposition via conjugation performed by
684 triparental mating as described previously[48]. Briefly, the transposon and helper plasmids were
685 delivered to the recipient from WM6026 donor strains by overnight incubation on LB agar plates
686 supplemented with 300 µM DAP and 0.5% glucose to repress AG expression. The next day,
687 cells were scraped and resuspended in 1 ml LB, and 1000X-20000X dilutions of this conjugation
688 mixture were plated on LB plates (lacking DAP) supplemented with the appropriate antibiotics
689 and 0.5% glucose. All strains were maintained under strict glucose repression of the pLlacO-1
690 promoter at every step until screens were performed. Thus, all AGs were maintained as single-
691 copy chromosomal inserts at the same attTn7 site in the genomic region between *glmS* and
692 *pstS* genes.
693 Efficient Tn7 transposition is necessary to construct an AG library in any given *E. coli*
694 host. Tn7 conjugation-transposition efficiencies were determined in 44 ECOR stains containing
695 an intact *glmS-pstS* region (we note that other strains can support Tn7 integration as well). Tn7
696 transposons carried only a barcode sequence and the gentamicin-resistance gene to minimize
697 payload-dependent variation in conjugation efficiency between strains. Efficiency of conjugation
698 was determined by streaking the conjugation mixture on two different gentamicin concentrations
699 (25 µg/ml and 10 µg/ml) to account for possible differences in natural antibiotic resistance
700 between the wild strains. The barcode sequence was verified by Sanger sequencing and both
701 flanks were checked by PCR to ensure proper integration of the Tn7 transposon at the expected
702 attTn7 site. Strains that supported Tn7 transposition at high levels and were amenable to agar-
703 overlay plaque assays were selected for further experiments.
704
705 **DNA extraction + Nextera Genomic Sequencing + Analysis**
706 Genomic DNA was extracted using a modified SDS–proteinase K method: Briefly, cells
707 pelleted from 50 µl of saturated culture (or 5 µl of the cell mixture scraped from selective plates
708 during AG library construction) were re-suspended in 200 µl of lysis buffer (10 mM tris, 10 mM
709 EDTA, 400 mg/ml proteinase K, 0.5% SDS) and incubated at 50-55°C for 1 hour. Subsequently,
710 the temperature was lowered to 37°C and RNase A (Thermo Scientific) was added to a
711 concentration of 1 mg/ml. After 30-60 min of incubation, the digests subsequently purified using

712 the Genomic DNA Clean & Concentrator Kit (Zymo Research). DNA was prepared for whole-
713 genome sequencing using the Nextera Flex Library Prep Kit (now, the Illumina DNA Prep Kit)
714 according to manufacturer's instructions, except that we lowered the volumes of all
715 tagmentation reactions 5-fold, and used a custom dual-indexed primer set JSW-SS-22:33
716 (CAAGCAGAAGACGGCATACGAGAT NNNNNNNN GTCTCGTGGGCTCGG) and JSW-SS-
717 34:41 (AATGATACGGCGACCACCGAGATCTACAC NNNNNNNN TCGTCGGCAGCGTC) to
718 amplify libraries for 11-13 cycles using Phusion High Fidelity PCR Master Mix (NEB) instead of
719 the Illumina-supplied PCR reagents. The N_8 sequences correspond to reverse-complemented
720 Nextera DNA indexes N701 to N712 and N501 to N508, respectively (Illumina). Libraries were
721 resolved by agarose gel electrophoresis and DNA was excised in the 300-400 bp range. Gel
722 slices were purified using the Zymo Gel DNA Recovery Kit and sequenced on the Illumina
723 MiSeq using 150-cycle v3 or 600-cycle v2 reagent kits.
724 Sequencing adaptors were trimmed from reads using cutadapt, mapped to reference
725 genomes using bowtie2, and visualized using IGV. Reference genomic contigs of ECOR strains
726 were obtained from the USDA[49]. SNPs, indels, and large structural variants were analysed
727 using a mapping-free variant finder[50].
728
729 **Phage propagation and plaque assays**
730 High-titer phage stocks were generated by growth in liquid culture. All phage infections
731 were performed at 37°C in LB supplemented with 10 mM $MgSO_4$ and 5 mM $CaCl_2$. *E. coli*
732 BW25113 or C990 were grown to an optical density (at 600 nm; OD_{600}) ~ 0.5 (1-5 X 10^8 cfu/ml)
733 and infected with the desired phage at MOI ~ 5. Infections were allowed to proceed at 37°C with
734 agitation for 5-6 hours until the cultures were clear. Lysates were clarified by centrifugation
735 (8000 X g for 2 min) and sterile-filtered using a 0.45 µm SFCA syringe filter (Millipore). Phage
736 lysates were stored at 4°C.
737 Bacteriophages T2, T3, T4, T5, T6, T7, P1vir were obtained from Dr. Vivek Mutalik
738 (LBNL), and λvir was obtained from Dr. Ry Young (Texas A&M). To the extent possible, all
739 phages were grown on *E. coli* C990 which lacks restriction-modification. T4 and P1vir could not
740 be grown on C990 and were instead grown on *E. coli* BW25113. To ensure that our phage
741 lysates were free from ancestral Type I R-M methylation marks or other host-passage derived
742 DNA modifications, we serially passaged all phages in the desired hosts 3 times. All phage
743 genomes were subsequently verified by Nextera high-throughput sequencing.
744 Phage titers were determined against various hosts by the agar-overlay method with LB
745 top agar containing 0.7% agar, 10 mM $MgSO_4$ and 5 mM $CaCl_2$. Phage spot-titration assays
746 were performed with 10-fold serial dilutions of phage lysates (3-5 µl of each dilution pipetted
747 onto bacteria immobilized in top-agar overlay using a multichannel pipette).
748
749 **AG Library Preparation + NGS verification + testing post-conjugation evenness**
750 200 AGs with unique 10 bp DNA barcodes were synthesized (Twist and IDT) and cloned
751 individually under pLlacO-1 control (IPTG inducible, repressed by glucose) into a modified
752 pTn7C185 vector (with mobile CRISPRi components sgRNA and dCas9 removed). Plasmid
753 assembly reactions were performed in 96-well format using the 2X Gibson assembly master mix
754 (NEB) and transformed into chemically-competent BW25141 cells in 96-well format.
755 Transformations were individually plated on LB agar supplemented with 20 µg/ml gentamicin
756 and 0.5% glucose and one colony was initially checked by Sanger sequencing from each plate.
757 Where a successful transformant could not be obtained on the first attempt, more colonies were
758 sequenced in successive rounds until a plasmid with no mutations in the region spanning the
759 pLlacO-1 promoter, the AG ORF, and the barcode sequence was obtained. Verified vectors
760 were miniprepped using the ZR Plasmid miniprep – Classic kit (Zymo) and transformed into
761 chemically-competent WM6026 cells in 96-well format, and each transformation was plated

individually on LB agar supplemented with 50 µg/ml gentamicin, 300 µM DAP, and 0.5% glucose.

One colony from each of the 200 WM6026 transformants was grown in 2 ml square-well 96-well plates with shaking overnight, and 100 ul of each culture was then pooled together to yield ~20 ml of the donor mixture. This mixture was concentrated down to 10 ml by centrifugation at 6000 g for 10 min, and 5 ml of 60% glycerol (20% final) was added to yield the donor library. The 15 ml library was divided into single-use 50 ul aliquots (each containing >50M cells, i.e. >300K of each donor strain) and stored at -80°C.

Genomic DNA was extracted from one of the library aliquots and all constructs in the final pool were verified by Nextera high-throughput sequencing. This allowed us to assess library evenness as well as verify the fidelity of our cloning, colony picking, and strain pooling process. Overall, 195 AGs were cloned successfully. One AG (*orf88*) never yielded any colonies in multiple cloning attempts, possibly due to its cytotoxicity. Three AGs (*orf36, orf52, orf94*) were dropouts during the WM6026 growth step, and another two (*orf25, orf84*) were discovered to contain mutations upon high-throughput sequencing that had been missed by Sanger sequencing. *Orf25* had previously presented difficulties in cloning, and did not yield a mutation-free construct even after checking >12 colonies (whereas most transformations yielded a correct construct on the first attempt). The *orf84* construct contained a mixed base at the last position of the second repeat of the Lac-operator sequence upstream of the ORF, which would not be expected to affect any downstream assays significantly.

The donor library was conjugated into MG1655 *E. coli* according to the Tn7 transposition procedure (above) and a 1000X dilution of the conjugation mixture was plated on LB agar supplemented with 20 µg/ml gentamicin and 0.5% glucose. We routinely use multiple 150 mm X 15 mm bacteriological petri dishes (Corning) to obtain sufficient numbers of transconjugant colonies to assemble libraries. The next day, >>2,000-20,000 transconjugant colonies (10-100X the number of AGs in the library) were scraped and combined thoroughly, and the thick cell mixture was directly frozen at -80°C in 20% glycerol. Genomic DNA was extracted from this mixture and the distribution of DNA barcodes corresponding to individual AGs was measured in both this post-conjugation strain mixture as well as the genomic DNA extracted from the WM6026 donor library aliquot.

To amplify barcodes for high-throughput sequencing, genomic DNA was subjected to two successive rounds of PCR. The DNA barcodes were amplified from genomic DNA using universal primers JSW-SS-170 (CGACGCTCTTCCGATCTNNNNN <u>TGATGTCGTTGTTGCCATCG</u>) and JSW-SS-171 (ACTGACGCTAGTGCATCA <u>CTTTCTGAGCCAGTGTTGCT</u>) and the Q5 Hot Start High-Fidelity 2X PCR master mix (NEB) according to manufacturer's instructions. Sequencing adaptors were attached in a second round of PCR using amplicons from the first PCR round as the templates, with dual-indexed primer sets JSW-SS-42:53 (CAAGCAGAAGACGGCATACGAGAT NNNNNNNN GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT <u>ACTGACGCTAGTGCATCA</u>) and JSW-SS-54:61 (AATGATACGGCGACCACCGAGATCTACAC NNNNNNNN ACACTCTTTCCCTACA<u>CGACGCTCTTCCGATCT</u>), where the N_8 sequences correspond to reverse-complemented TruSeq HT indexes D701 to D712 and D501 to D508, respectively (Illumina). Template-matching regions in the primers are underlined. Cycling conditions for round 1 were as follows: one cycle at 98°C for 1 min; two cycles at 98°C for 10s, 66°C for 30s, and 72°C for 10s; 22 cycles at 98°C for 10s, and 72°C for 20s; and one cycle at 72°C for 2 min. Conditions for round 2 were one cycle at 98°C for 1 min; two cycles at 98°C for 10 s, 64°C for 30 s, and 72°C for 10 s; 4 cycles at 98°C for 10 s, and 72°C for 20 s; and one cycle at 72°C for 2 min. 0.5-1 ul of unpurified 1st round reaction product was used as template for the 2nd round of PCR. Amplicons from the 2nd round were gel-purified by electrophoresis (3% agarose gel, 4.2 V/cm, 2 hours) and quantified using the 1X dsDNA HS kit with the Qubit 4.0 Fluorometer (Invitrogen). Amplicons were sequenced on a MiSeq using the 150-cycle v3 reagent kit

813 (Illumina) in single-end format (100 bp) with two (8 bp) index reads. DNA barcodes were
814 trimmed from the reads and the proportions of various AGs in the mixtures were determined
815 using the prevalence of the corresponding barcode sequence.
816 Concordance between the Nextera whole-genome sequencing and the barcode
817 sequencing of the WM6026 donor mixture showed that barcode representation was a suitable
818 proxy for the true distribution of AGs in the mixture and that amplification biases during library
819 preparation were not a significant source of noise. Concordance between barcode sequencing
820 in the WM6026 donor and MG1655 post-conjugation recipient mixtures indicated that the Tn7
821 transposition and colony scraping process does not meaningfully bottleneck AG distribution.
822 The AG library was then conjugated (in triplicate) into ECOR strains that were previously
823 confirmed to support robust Tn7 transposition. Strain mixtures were prepared in a procedure
824 analogous to library construction in MG1655, except higher dilutions of the ECOR+AG
825 conjugation mixtures (10,000-20,000X) were plated and the gentamicin concentration on
826 selective plates was lowered to 10 µg/ml to account for higher steady state cell densities and
827 higher gentamicin sensitivity of the wild *E. coli* strains. As with the MG1655 library, thick cell
828 mixtures comprised of >>2,000-20,000 colonies scraped from each replicate conjugation were
829 directly stored at -80°C in 20% glycerol and used to seed cultures for the AG screens (below).
830 Genomic DNA was extracted from each mixture and barcode sequencing was performed to
831 measure the AG distribution in every replicate of all ECOR strain derived libraries. The
832 replicates were very highly correlated, and AG distributions in all libraries were confirmed to be
833 free of significant bottlenecking. These datasets are the reference libraries that serve as the
834 baseline for AG induction and phage infection experiments.
835
836 **MOI calculation for liquid killing assays**
837 Phage adsorption to host cells may be inefficient in very dilute cultures (10^{6-7} cfu/ml).
838 Therefore, the apparent multiplicity of infection (MOI-a) might be lower than the expected MOI
839 (MOI-e). We added 2x serial dilutions of a λvir lysate to a 1:100 dilution of a saturated culture of
840 *E. coli* BW25113 that had been allowed to acclimate at 37°C with shaking for 15 min. The
841 proportion of surviving cells was determined (relative to uninfected controls) after a 60-75 min
842 incubation (one phage replication cycle) by plating on LB agar. An apparent MOI value was
843 calculated using the Poisson distribution function with zero occurrences $MOIa = -ln\,(MOIe)$ for
844 each phage dilution, assuming that all infected cells would be killed by the virus (i.e. the
845 proportion of surviving cells would be the zero class in the Poisson distribution). The apparent
846 and expected MOI values were plotted against each other and the inverse of the slope (1/m) of
847 the line (MOI-a = m*MOI-e + 0) fit through the data was determined as the fold-change
848 reduction in MOI-a as compared to MOI-e for these experimental conditions. This fold-change
849 reduction was determined to be ~26x. Hence, all phage infection screens were targeted to an
850 MOI-a value of ~100 to ensure efficient phage adsorption and infection within the timeframe of a
851 single phage replication cycle.
852
853 **AG Screen + Plaque assay confirmation + Data analysis**
854 Thick cell mixtures from each of the three replicates of the various ECOR+AG libraries
855 were thawed on ice and diluted into LB supplemented with Mg^{2+}, Ca^{2+}, and 10 µg/ml gentamicin.
856 Typically, 1-5 µl of the library mixture was diluted in 2 ml broth to yield an initial OD_{600} of 0.1-
857 0.15. The cultures were immediately split x2, with one tube used to monitor OD_{600} values as the
858 experiments progressed, and the other left untouched for the screen. AG expression was
859 induced by the addition of 1 mM IPTG after acclimatization for 15 min at 37°C in a shaking
860 incubator and OD_{600} values were measured again for each ECOR+AG library at time of
861 induction. Libraries were then grown for ~80-120 min depending on the relative growth rates of
862 the various ECOR strains, to OD_{600} ~ 1.2-1.5. A 10^{-5} dilution of each late-log phase culture was
863 then plated on LB agar (with 10 µg/ml gentamicin and 0.5% glucose to repress AGs and avoid

fitness costs) to obtain a count of viable colony forming units (cfu). Cultures were immediately diluted ~1:20 in fresh LB (with IPTG, Mg^{2+}, Ca^{2+}, and 10 µg/ml gentamicin) and 1 ml aliquots were gently mixed with phage lysates at an MOI-e of ~30-100 (the actual MOI-e values were recorded the next day when cfu counts were obtained). Parallel uninfected controls were maintained for each experiment, even when the same ECOR+AG library was being tested against different phages and the uninfected controls would be interchangeable. Infections were allowed to proceed for 30-35 min for T-phages and 60-70 min for P1vir and λvir, i.e. one replication cycle. AG induction was limited to 1-2 hrs and subsequent infections were only allowed to proceed for ~1 phage replication cycle to prevent prolonged induction of various AGs from biasing the results. This setup also avoids biases resulting from epigenetic modification of phages by widespread restriction-modification (R-M) defenses in *E. coli*. However, we note that allowing sufficient time for the phages to complete cell lysis was necessary as fast-growing cells were sometimes able to form colonies on LB agar despite having been infected. These colonies had an uneven size distribution and a “splotchy” appearance, perhaps due to ongoing infection on the plate that was nevertheless not able to outpace colony formation. After the infection period was complete, the cultures were vortexed, diluted 1:100, vortexed again, and 50 µl of the dilution was plated on 2 large (150 mm X 15 mm) LB agar plates (with 10 µg/ml gentamicin and 0.5% glucose to repress AGs and avoid fitness costs). The next day >>2,000-20,000 colonies were scraped from each infected and uninfected set of plates. 5 µl of the thick resulting cell mixtures was reserved for genomic DNA extraction, and the rest was frozen at -80°C in 20% glycerol as a backup. A full technical record of our screens is provided in Supplementary File 4.

AG barcodes were amplified and sequenced as with the WM6026 donor library, with a slight modification. The three biological replicate sets of ECOR+AG experiments were amplified using JSW-SS-171 and different versions of primer JSW-SS-170 containing extra N’s at the start of the sequencing read. The first replicate was amplified with JSW-SS-170 (containing 5 N’s), the second replicate with JBD-SS-254 (CGACGCTCTTCCGATCTNNNNNN TGATGTCGTTGTTGCCATCG) with 6 N’s, and the third replicate with JBD-SS-255 (CGACGCTCTTCCGATCTNNNNNNN TGATGTCGTTGTTGCCATCG) with 7 N’s. This ensured that reads from the three sets would not cross contaminate analyses of replicate experiments. Amplicons were sequenced to the same depth as the baseline libraries on Illumina MiSeq and NextSeq instruments in single-end format (100 bp) with two (8 bp) index reads.

We required exact matches for the 20 bp primer-derived invariant sequences flanking the barcodes. These sequences were trimmed and the proportions of various AG-encoding strains in the mixtures were determined using the prevalence of the corresponding 10 bp barcode sequences. Barcode counts per AG were normalized by dividing each count by the median of each experiment, which would represent neutral AG fitness. A log2 transformation was subsequently applied to obtain fitness scores. Care was taken to sequence each library to similar depth such that fitness scores across experiments were distributed in the same range. Poorly represented AGs were identified in the baseline libraries as well as the infected and uninfected experimental samples using a low read cutoff of 100, or if the log2 normalized average read counts across the three replicates for any AG were >3x the median absolute deviation below the median for that experiment. Reproducibility was determined using T-tests across the three replicates, and we enforced a P-value cutoff of 0.2. The magnitude of the AG fitness effect was determined by (a) subtracting the AG fitness scores in the uninfected library from the infected libraries to assess whether AGs changed host sensitivity to phage infection, and (b) subtracting the AG fitness scores in the baseline library from the uninfected libraries to assess whether AGs change host fitness directly. AGs were selected for further study if they were not poorly represented to begin with, produced reproducible effects across replicates, and changed host fitness by at least 4-fold. Although we had targeted our experiments to phage-host combinations where the host was mostly resistant to the phage in order to find counter-defense AGs, a few hosts turned out to be more sensitive to certain phages in liquid infection

915 than suggested by agar-overlay plaque assays (perhaps due to abortive antiviral defenses). In
916 these cases, our platform was also able to robustly identify AGs that conferred a protective
917 advantage to the host (e.g. Imm and Cor, which both block phage DNA entry [51,52]). In practice
918 however, these effects can be very variable and the reproducibility filters typically needed to be
919 relaxed in order to find such superinfection exclusion AGs.
920 All counter-defense AGs were subsequently conjugated into the wild ECOR hosts where
921 they demonstrated an effect, and tested against all 8 phages (not just the ones that were used
922 in the screen with that host). In several hosts, the AG was found to additionally sensitize the
923 host to phages that were omitted in the screen. Not all the screen phenotypes could be
924 confirmed in agar-overlay plaque assays, possibly due to variations in cell growth between liquid
925 and solid media, or due to the requirement for successive infection cycles for plaque formation
926 that would render plaque assays less sensitive than single-infection cycle dynamics in our
927 screen. Nevertheless, further experiments were performed only in phage-host-AG combinations
928 that could be confirmed by plaque assay due to ease of follow up testing.
929 Conditional lethal AGs were identified directly from the heatmap of fitness scores, with a
930 subset selected for further experimentation using an arbitrary fitness cutoff of ~6-7 (~100-fold).
931
932 **Tn5 Dropout Screen + Data analysis to find phage exclusion systems**
933 Tn5 whole-genome knockout strain libraries were constructed using the RB-TnSeq
934 donor library (pKMW7-derived strain APA766 from Dr. Adam Deutschbauer) conjugated with the
935 recipient ECOR strain in a biparental mating on LB plates supplements with 300 μM DAP. The
936 next day, cell mixtures were scraped and resuspended in 1 ml LB and 1000X dilutions of this
937 conjugation mixture were plated on 10 large (150 mm X 15 mm) LB plates (lacking DAP)
938 supplemented with 50 μg/ml kanamycin. Approximately 0.5-1 million colonies were scraped
939 taking care to break colonies apart on the surface of the plates, mixed very thoroughly,
940 aliquoted and stored at -80°C with 20% glycerol, with ~5-10 μl reserved for genomic DNA
941 extraction and reference library preparation.
942 RB-TnSeq allows the same mutant library to be challenged under various conditions
943 such as challenge by various phages, and simplifies the downstream sequencing-library
944 preparation by substituting barcode sequencing for all experimental conditions after initial
945 characterization of the reference library. To characterize the Tn5-knockout reference libraries in
946 the various ECOR strains, genomic DNA was extracted from 5-10 μl of the cell scrape and
947 prepared for sequencing using the NEBNext Ultra II FS DNA Library Prep Kit (NEB) with the
948 following modifications. 500 ng DNA was fragmented for 25 min according to manufacturer's
949 instructions. The adaptor ligation step was performed as directed albeit with a custom pre-
950 annealed oligo adaptor comprising of YL001 (/5Phos/GATCGGAAGAG/3ddC/) annealed to one
951 of YL002:005
952 (AGCGGCAATTTCACACAGGACAAGCAGAAGACGGCATACGAGATNNNNNNNNAGTGGTG
953 ACTGGAGTTCAGACGTGTGCTCTTCCGATC*T) where the 4-base underlined sequence
954 serves as a variable barcode unique to each oligo. The USER Enzyme steps post-ligation were
955 skipped and the bead-cleanup was performed using 40 μl and 20 μl of Sample Purification
956 Beads successively as instructed for library size selection in the 150-250 bp range. The PCR
957 amplification of transposon-genome junctions was performed using the cycling parameters
958 described in the kit for 11 cycles with Q5 Ultra II FS Master Mix (NEB) using primers YL006
959 (AGCGGCAATTTCACACAGGA) and JBD-SS-272/oAD493
960 (ACACTGGCAGAGCATTACGCCCT) instead of the kit-supplied primers. A second nested-PCR
961 enrichment step was performed for 11 cycles using the same cycling parameters as before,
962 using the entire bead-purified reaction product from the first step as a template with primers
963 YL009 (CAAGCAGAAGACGGCATACGAG) and JBD-SS-273:280
964 (AATGATACGGCGACCACCGAGATCTACAC TATAGCCT
965 ACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNNNN

966 GCAGGGATGTCCACGAGGTCTCT) where N_8 barcodes correspond to reverse-complemented
967 Illumina TruSeq HT D501-508 indexes. Bead-purified amplicons were quantified using the 1X
968 dsDNA HS kit with the Qubit 4.0 Fluorometer (Invitrogen) and sequenced on Illumina MiSeq and
969 NovaSeq instruments in single-end format (150 bp) with two (8 bp) index reads.
970 The first 100 bp of these reads contain transposon derived sequences and a DNA
971 barcode that uniquely identifies a particular Tn5 insertion. The last 50 bp of these reads contain
972 host-genome derived sequence that help determine the location of the transposon insertion and
973 hence construct a map between the Tn5 barcode and the insertion site (IS) in the genome.
974 Briefly, 20 bp barcodes and 16-50 bp IS sequences were trimmed from the sequencing reads.
975 To isolate the barcode and IS, we allowed 3 mismatches in the 23 bp invariant sequence
976 upstream and up to 6 mismatches in the 50 bp downstream of the barcode using a Hamming
977 distance function to compare the expected and actual sequences. Any reads failing these
978 sequence-fidelity requirements were discarded. Reads containing an IS that matched the Tn5
979 transposon vector in the donor strain were also discarded. The 4 Ns adjacent to the underlined
980 index sequences in YL002:005 were also sequenced as part of one 8 bp index read and served
981 as a tag to remove PCR duplicates. This allowed us to distinguish between sequencing "reads"
982 and DNA "fragments" present in the sample at the adaptor ligation step. Thus, reads containing
983 unique combinations of barcode, PCR tag, and IS were entered into the barcode/IS map for
984 each Tn5 library, each representing a distinct DNA fragment (but also machine errors in
985 sequencing, chimeric PCR amplicons, 3' truncations etc.). The filtered ISs were then mapped to
986 genes within their respective host genomes. As described previously[53], it is possible for
987 transposon barcodes to map to the genomes ambiguously. Barcodes associated with multiple
988 distinct ISs were retained only if their most frequently represented IS accounted for over 90% of
989 the fragments containing that barcode. We then counted the number of DNA fragments
990 recovered per gene and used this as a reference library for subsequent BarSeq experiments to
991 assay depletion of strains with mutations in specific genes upon phage infection.
992 Tn5 mutant strain libraries of the various ECOR strains were then challenged by various
993 phages in a procedure analogous to the ECOR+AG screens. Cultures of Tn5 libraries were
994 initiated with 100 µl of thick cell scrape mixture in 100 ml LB supplemented with 50 µg/ml
995 kanamycin, Mg^{2+}, and Ca^{2+}, to an initial OD_{600} value of 0.1-0.15. Cultures were grown at 37°C
996 with shaking for 80-90 min to an OD_{600} of 0.7-1. A 10^{-5} dilution of each late-log phase culture
997 was then plated on LB agar (with 50 µg/ml kanamycin) to obtain a count of viable colony forming
998 units (cfu). Cultures were immediately diluted ~1:10 in fresh LB (with Mg^{2+}, Ca^{2+}, and 50 µg/ml
999 kanamycin) and 1 ml aliquots were gently mixed with lysates of the appropriate phages at an
1000 MOI-e of ~30-100 (the actual MOI-e values were recorded the next day when cfu counts were
1001 obtained). Since the parental ECOR strains were typically quite resistant to phage infection to
1002 begin with, only mutants with transposon-inactivated defense systems would be expected to
1003 drop out upon infection. In practice however, some strains exhibited more pronounced mortality
1004 upon prolonged phage infection in the Tn5 screens. In particular, in experiments with T2 and the
1005 Tn5 library constructed in the Δ*wecA* strain of ECOR21 (which was generally more susceptible
1006 to infection by T2 and other phages), MOI was titrated in three parallel infection screens (with
1007 varying amounts of phage lysate added: 350 µl, 100 µl, and 25 µl) to empirically ensure specific
1008 depletion of only the phage-sensitive mutants. Parallel uninfected controls were maintained for
1009 each experiment, and each set of experiments was performed in biological duplicates. Infections
1010 were allowed to proceed for 5 min at the benchtop without shaking, and 75 min at 37°C with
1011 shaking. Since an interaction between the host and phage had already been identified in the
1012 earlier AG screen, infections were not restriction to a single phage replication cycle, but allowed
1013 to proceed further to achieve full depletion of the phage-sensitive Tn5-knockout mutants in the
1014 pool. After the infection period was complete, the cultures were vortexed, diluted 1:10, vortexed
1015 again, and 100 µl of the dilution was plated on 2 large (150 mm X 15 mm) LB agar plates (with
1016 50 µg/ml kanamycin). We note that this dilution yields bacterial lawns and in our experience this

1017 preserves diversity of surviving colonies without degrading the signal of dropouts by phage-
1018 killing in the experiment. The next day, bacterial lawns were scraped from each infected and
1019 uninfected set of plates. 5 µl of the thick resulting cell mixtures was reserved for genomic DNA
1020 extraction, and the rest was frozen at -80°C in 20% glycerol as a backup. A full technical record
1021 of our screens is provided in Supplementary File 8.
1022 To amplify barcodes from the post-selection infected and uninfected Tn5 libraries, we
1023 followed a modified BarSeq98 method[53], with two successive rounds of PCR. The transposon
1024 barcode sequence was amplified by PCR using Q5 Ultra II FS Master Mix (NEB) with primers
1025 JBD-SS-310 (CGACGCTCTTCCGATCTNNNNNNN GATGTCCACGAGGTCTCT) and JBD-SS-
1026 311 (ACTGACGCTAGTGCATCA GTCGACCTGCAGCGTACG). Template-matching regions in
1027 the primers are underlined. Sequencing adaptors were attached in a second round of PCR
1028 using amplicons from the first PCR round as the templates, with dual-indexed primer sets JSW-
1029 SS-42:53 and JSW-SS-54:61 (sequences already provided in section describing AG screening).
1030 Cycling conditions for round 1 were as follows: one cycle at 98°C for 4 min; two cycles at 98°C
1031 for 30s, 63°C for 30s, and 72°C for 30s; 22 cycles at 98°C for 30s, and 72°C for 30s; and one
1032 cycle at 72°C for 4 min. Conditions for round 2 were one cycle at 98°C for 1 min; two cycles at
1033 98°C for 10 s, 64°C for 30 s, and 72°C for 10 s; 4 cycles at 98°C for 10 s, and 72°C for 20 s;
1034 and one cycle at 72°C for 2 min. 0.5-1 ul of unpurified 1st round reaction product was used as
1035 template for the 2nd round of PCR. Amplicons from the 2nd round were gel-purified by
1036 electrophoresis (3% agarose gel, 4.2 V/cm, 2 hours) and quantified using the 1X dsDNA HS kit
1037 with the Qubit 4.0 Fluorometer (Invitrogen). Amplicons were sequenced on Illumina MiSeq and
1038 NovaSeq platforms in single-end format (150 bp) with two (8 bp) index reads.
1039 DNA barcodes were trimmed from the reads and the proportions of various Tn5-
1040 knockout mutants in the mixtures were determined using the prevalence of the corresponding
1041 barcode sequence. Briefly, we allowed 3 mismatches in the 18 bp invariant sequence upstream,
1042 and 4 mismatches in the 18 bp downstream of the transposon barcodes using a Hamming
1043 distance function to compare the expected and actual sequences. The frequency of occurrence
1044 of each barcode (representation number) was counted. Transposon barcodes were matched to
1045 those found in the reference libraries, inferred to represent knockouts in the genes containing
1046 the corresponding ISs, and the various barcode representation numbers were added together
1047 for each gene. Cumulative transposon insertions per gene were then compared between the
1048 infected and uninfected samples (with the reference library as a baseline to filter out poorly
1049 represented, presumably essential genes) using the same data analysis procedure as for the
1050 AG screen, albeit with two biological replicates instead of three.
1051
1052 **Tn5 Suppressor Screen + Data analysis to find abortive loci**
1053 Tn5 whole-genome knockout strain libraries were constructed using the RB-TnSeq
1054 donor conjugated with the recipient ECOR strain as before, with modifications. No phages are
1055 involved. The recipient ECOR strain was equipped with the AG (as a single copy chromosomal
1056 Tn7 insertion as usual) which triggers PCD, presumably by activating an Abi system. The trigger
1057 AG was inserted prior to Tn5 library construction in order to minimize passages of the library
1058 prior to screening and avoid potential bottlenecking due to fitness effects of the various Tn5
1059 knockouts. Libraries were constructed as above, except that conjugation was carried out under
1060 strict glucose repression to prevent PCD activation at this stage, and the conjugation mixture
1061 scraped from the LB+DAP+glucose plates was then plated directly on LB plates containing 10
1062 µg/ml gentamicin and 1 mM IPTG in addition to 50 µg/ml kanamycin. We typically needed to
1063 plate the entire conjugation mixture scraped from five 100 mm LB+DAP+glucose plate onto five
1064 150 mm LB+IPTG+Kan+Gm plate in order to obtain sufficient numbers of suppressor colonies
1065 for characterization by high-throughput sequencing. Control samples were obtained by plating
1066 1000-2000X dilutions on LB+glucose+Kan+Gm plates. Suitable dilutions were empirically
1067 determined for the various ECOR hosts. The next day, suppressor colonies were scraped from

1068 the large plates and stored at -80°C with 20% glycerol, with ~5-10 µl reserved for genomic DNA
1069 extraction and sequencing. Genomic DNA was extracted and libraries were prepared and
1070 sequenced akin to the treatment of the RB-TnSeq/NEBNext reference libraries detailed above.
1071 There was no need for barcode sequencing since the screen identifies survivors instead of
1072 dropouts, and each library was only characterized once after IPTG induction/selection for
1073 suppressor mutants. Tn5 reference libraries were constructed from the data using the same
1074 procedures for the baseline libraries constructed in the various naïve (without AGs) ECOR
1075 strains above. The number of transposon insertions (i.e. the number of uniquely mapping ISs
1076 from distinct DNA fragments) recovered per gene were used directly as a measure of survival of
1077 mutants despite expression of PCD-triggering AGs. In practice, no statistical analysis was
1078 necessary to identify the genetic loci responsible for PCD, as the signal-to-noise ratio was
1079 sufficiently high to allow visual identification of hits when transposon insertion frequency was
1080 plotted in control vs experimental samples (i.e. when the trigger AG was repressed by plating on
1081 glucose-containing media, vs induced with IPTG).
1082
1083 **LPS gel electrophoresis**
1084 Lipopolysaccharide was prepared from liquid overnight cultures of wild *E. coli* strains
1085 expressing the relevant AGs (LB + IPTG). Briefly, 50-100 µl of saturated culture was spun down
1086 and cells were thoroughly resuspended in 100-200 µl of lysing buffer (62.5 mM Tris-HCl, pH6.8,
1087 10% glycerol, 2% SDS, 4% ß-mercaptoethanol). We routinely supplement 1X Laemmli Buffer
1088 (BioRad) with SDS and ßME to make lysing buffer. Samples were boiled at 100°C for 15 min,
1089 and allowed to rest at room temperature for 15 min. 100 µg of Protease K were added to each
1090 sample and protein was digested at 60°C for at least 1 hr. 15 µl of each sample were run onto a
1091 12% polyacrylamide TGX Mini-Protean Precast gel (BioRad) at 200V for 35 min, and visualized
1092 using the Pro-Q Emerald 300 LPS stain kit (Thermo Fisher) precisely according to
1093 manufacturer's instructions.
1094
1095 **Allelic Exchange for *wecA* deletion**
1096 The R6K-origin pKEK2201 vector (from Prof. Karl E. Klose) carrying the desired genetic
1097 modification (eg. a *wecA* deletion) flanked by 1000 bp homology arms was prepared by Gibson
1098 assembly, and delivered as a suicide vector into wild *E. coli* strains via conjugation from the
1099 WM6026 auxotrophic donor. Conjugation was performed using the same method used to deliver
1100 AGs, albeit with biparental instead of triparental mating. Successful transconjugants (containing
1101 a chromosomally integrated plasmid after the first homologous recombination event HR1) were
1102 selected by incubation on LB plates supplemented with kanamycin. The second recombination
1103 event HR2 to complete the allelic exchange was selected via *sacB*-mediated counterselection
1104 on LB plates supplemented with 300 mM sucrose after overnight growth of HR1
1105 transconjugants. Successful allele exchange was verified by PCR, and by LPS electrophoresis
1106 to assay for the absence of O-antigen in the case of the *wecA* deletion.
1107
1108 **Live cell Microscopy**
1109 Cells were prepared for microscopy, immobilized under an LB agarose pad, and imaged
1110 as described previously[54]. Where necessary, AGs were pre-induced with 1 mM IPTG for 30 min
1111 immediately before harvesting for microscopy.
1112
1113 **Verification of abortive loci activity**
1114 Abortive/PCD systems identified by transposon screening of conditionally lethal AGs
1115 were cloned by Gibson assembly using the NEBuilder HiFi DNA assembly master mix (NEB)
1116 onto pBAD Myc/His A vectors along with their native promoters wherever possible. AGs that
1117 trigger PCD systems (or appropriate controls) were conjugated into DH10ß/TOP10 strains
1118 carrying pBAD-PCD plasmids using the Tn7 transposition method (above). Where the

1119 restriction/PCD systems prevented conjugation of AGs, the order of transformations was
1120 reversed – the AG was conjugated into naïve DH10ß/TOP10 cells first, and subsequently these
1121 strains were transformed with the pBAD-PCD vectors. Successful transconjugants were grown
1122 overnight with glucose to repress trigger AGs, and viability was tested the next day by spotting
1123 10-fold serial dilutions of saturated cultures on LB agar plates with four additives: 1) Glucose
1124 repression to assess background toxicity, (2) Arabinose induction of PCD systems alone to
1125 assess fitness costs of over-expression of abortive components without the trigger AGs, (3)
1126 IPTG to assess ability of AGs to trigger PCD without over-expression of abortive components
1127 (i.e. under leaky expression conditions, or when expressed from the native promoters where
1128 present), and (4) IPTG and Arabinose to assess ability of AGs to trigger PCD when abortive
1129 components were also over-expressed.
1130
1131 **Affinity purification of AGs from wild *E. coli* hosts**
1132 FLAG-tagged AGs were conjugated into ECOR strains by Tn7-transposition (above)
1133 according to Supplementary File 9. Each strain was grown overnight at 37°C in 2 mL of LB
1134 supplemented with .5% glucose and 10 mg/mL gentamicin, then diluted 1:100 into 100 mL LB
1135 supplemented with 10 mg/mL gentamicin and grown at 37°C for 30 minutes. The media was
1136 then supplemented with 1 mM IPTG and grown for another 2.5 hours. "Popcorn" pellets were
1137 collected by centrifugation for 30 minutes at 4,000 X g, resuspension in 100 µL of ice-cooled
1138 lysis buffer [50 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM $MgCl_2$, 0.5% NP40
1139 (Research Products International), 125 U/mL Benzonase (Millipore), 1x protease inhibitor
1140 cocktail (Roche, cOmplete ULTRA tablets, mini, EDTA-free), 0.5 mg/ml lysozyme (Fisher
1141 Scientific)], followed by dropwise addition into liquid nitrogen and stored at -80°C. Popcorn
1142 pellets were lysed by 10 cycles of cryomilling for 2 minutes at 12 counts per second in a SPEX
1143 SamplePrep 6870D Freezer/Mill and stored at -80°C.
1144 For each phage-host combination, 3 replicate lysates were reconstituted using 200 mg
1145 of the cryomilled powder resuspended in 750 µL of lysis buffer. Lysates were cleared by
1146 centrifugation at 16,000 X g at 4°C for 15 minutes. Anti-FLAG magnetic agarose beads (Pierce)
1147 were washed on a magnet with 1 mL of ice-cooled IP buffer (50 mM Tris pH 7.4, 150 mM NaCl,
1148 1 mM EDTA, 1mM $MgCl_2$) twice and resuspended to their original volume in IP buffer. Cleared
1149 lysates were incubated with 30 µL of FLAG beads on an end-over-end tube rotator at 10 rpm at
1150 4°C overnight. The supernatant was separated from the beads on a magnet. Beads were
1151 washed once with 700 µL of .05% NP-40 and 1x protease inhibitor cocktail in IP, once with 700
1152 µL of .05% NP-40 in IP, and 3 times with 700 µL of IP. Proteins were eluted by incubation with
1153 25 µL of 100 µg/ml 3xFLAG peptide (Sigma) in 0.05% RapiGest (Waters Corp) in IP buffer on
1154 an end-over-end tube rotator at 10 rpm at 4°C overnight. The eluate was separated from the
1155 beads on a magnet. The elution step was then repeated with the same beads and fresh elution
1156 buffer at room temperature for 1 hour. The eluates were combined and stored at -20°C.
1157
1158 **Mass spectrometry sample preparation and data acquisition**
1159 Samples were prepared for Mass Spectrometry (MS) as described previously in the
1160 Filter-Aided Sample Preparation Protocol[55] with the following modifications. Before transferring
1161 to an AcroPrep Advance 10k Omega plate, each eluate was resuspended in 100 µL of 8M Urea
1162 (Thermo Scientific) dissolved in 50 mM ammonium bicarbonate (MP Biomedicals) and 5 mM
1163 TCEP (Aldrich), followed by centrifugation to dryness. During alkylation of cysteine residues, 10
1164 mM chloroacetamide (Thermo Scientific) was used instead of 10 mM iodoacetamide. Wash
1165 steps prior to proteolysis were carried out with 200 µL of 20 mM ammonium bicarbonate
1166 (instead of 100 µL). Proteolysis was performed without the addition of lysyl endopeptidase.
1167 Proteolysis was quenched by addition of 10% trifluoroacetic acid (Thermo Scientific) to reach ≈
1168 pH 2. Peptides were then purified over a BioPureSPE Mini 96-Well Plate (Nest Group). Briefly,
1169 C18 columns were activated by washing once with 400 µL of acetonitrile (Fisher Scientific) and

1170 twice with 400 µL of 0.1% formic acid (Thermo Scientific) in water solution, LC–MS grade
1171 (Thermo Scientific). Samples were loaded and centrifuged at 1500 X g for 2 minutes. The bound
1172 peptides were washed 3 times with 400 µL of 1% formic acid before elution with 150 µL of 50%
1173 acetonitrile in 0.1% trifluoroacetic acid. Peptides were vacuum dried and stored at -80°C.
1174 Samples were resuspended in 15 µl of 0.1% formic acid prior to loading for MS. The MS
1175 acquisition was performed as described[56] with same MS parameters and LC configuration.
1176
1177 **Mass Spectrometry data analysis and statistical analysis**
1178 Raw files were processed in FragPipe[57] using the LFQ-MBR workflow with minor
1179 modifications. The files were searched against a combined database of protein sequences from
1180 ECOR1, ECOR3, ECOR15, ECOR21, ECOR22, ECOR66[49], and MG1655 with duplicate entries
1181 removed and common contaminants added. Decoys were generated by pseudo-inversion.
1182 Protein and peptide FDR was fixed to 1% and MBR was disabled. Cysteine carbamylation was
1183 set as fixed modification while N-term acetylation, methionine oxidation and pyro-glu formation
1184 were set as variable modification with a maximum of 3 modifications per peptide.
1185 MaxLFQ intensities were used for statistical analysis[58]. Each triplicate set of host-AG
1186 samples was quantile normalized against a triplicate set of samples from the same host
1187 expressing *orf74*. For quantile normalization, all proteins in each sample were ranked according
1188 to their MaxLFQ intensity. Next, the mean value of the intensities of proteins occupying a given
1189 rank was calculated, then each protein's value was replaced with the mean value according to
1190 its rank. Fold changes were calculated by taking the log2 ratio of average normalized values
1191 between the test set and the control set. For each host-AG sample, all positively enriched
1192 proteins were ranked according to their fold change. For each AG, proteins were assigned
1193 "summed rank" values by calculating the sum of the ranks of each protein across all hosts
1194 where that AG was tested. P-values were calculated by student's t-test on the normalized sets
1195 of replicates from the test and control samples. Top enriched candidates for each AG were
1196 selected based on their summed ranks and p-values.
1197
1198 **Ronin systems phylogenetic analysis**
1199 The *ronA* nucleotide sequence in the ECOR55 Ronin system was used to identify similar
1200 genes by BLASTN. *RonA*-encoding bacterial genomes were downloaded using the entrez direct
1201 e-utilities command line tools. Of 100 complete genomes retrieved, *ronA* homologs had been
1202 annotated in 84. *hsdR* genes and other Type I R-M components were identified using phmmer
1203 and Defense Finder[59] respectively. Ronin-adjacent genetic architectures were visualized using
1204 Clinker[60]. All ronin-adjacent architectures are shown in HTML files with mouse-over annotations
1205 for genes.

1206 **Supplementary References**
1207

1208 46 Steinegger, M. & Soding, J. MMseqs2 enables sensitive protein sequence searching for
1209 the analysis of massive data sets. *Nat Biotechnol* **35**, 1026-1028 (2017).
1210 https://doi.org:10.1038/nbt.3988

1211 47 Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High
1212 throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.
1213 *Nat Commun* **9**, 5114 (2018). https://doi.org:10.1038/s41467-018-07641-9

1214 48 Peters, J. M. *et al.* Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi.
1215 *Nat Microbiol* **4**, 244-250 (2019). https://doi.org:10.1038/s41564-018-0327-z

1216 49 Patel, I. R. *et al.* Draft Genome Sequences of the Escherichia coli Reference (ECOR)
1217 Collection. *Microbiol Resour Announc* **7** (2018). https://doi.org:10.1128/MRA.01133-18

1218 50 Hansen, L. *SeqDiff*, <https://github.com/hansenlo/SeqDiff> (2021).

1219 51 Lu, M. J. & Henning, U. Superinfection exclusion by T-even-type coliphages. *Trends*
1220 *Microbiol* **2**, 137-139 (1994). https://doi.org:10.1016/0966-842x(94)90601-7

1221 52 Uc-Mass, A. *et al.* An orthologue of the cor gene is involved in the exclusion of
1222 temperate lambdoid phages. Evidence that Cor inactivates FhuA receptor functions.
1223 *Virology* **329**, 425-433 (2004). https://doi.org:10.1016/j.virol.2004.09.005

1224 53 Wetmore, K. M. *et al.* Rapid quantification of mutant fitness in diverse bacteria by
1225 sequencing randomly bar-coded transposons. *mBio* **6**, e00306-00315 (2015).
1226 https://doi.org:10.1128/mBio.00306-15

1227 54 Li, Y. *et al.* A family of novel immune systems targets early infection of nucleus-forming
1228 jumbo phages. *bioRxiv*, 2022.2009.2017.508391 (2022).
1229 https://doi.org:10.1101/2022.09.17.508391

1230 55 Fossati, A. *et al.* System-Wide Profiling of Protein Complexes Via Size Exclusion
1231 Chromatography-Mass Spectrometry (SEC-MS). *Methods in molecular biology* **2259**,
1232 269-294 (2021). https://doi.org:10.1007/978-1-0716-1178-4_18

1233 56 Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug
1234 repurposing. *Nature* **583**, 459-468 (2020). https://doi.org:10.1038/s41586-020-2286-9

1235 57 Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I.
1236 MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-
1237 based proteomics. *Nature methods* **14**, 513-520 (2017).
1238 https://doi.org:10.1038/nmeth.4256

1239 58 Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization
1240 and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-
1241 2526 (2014). https://doi.org:10.1074/mcp.M113.031591

59 Tesson, F. *et al.* Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat Commun* **13**, 2561 (2022). https://doi.org:10.1038/s41467-022-30269-9

60 Gilchrist, C. L. M. & Chooi, Y.-H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473-2475 (2021). https://doi.org:10.1093/bioinformatics/btab007

61 Loenen, W. A. & Murray, N. E. Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *Journal of molecular biology* **190**, 11-22 (1986). https://doi.org:10.1016/0022-2836(86)90071-9

62 Johnson, M. C. *et al.* Core defense hotspots within Pseudomonas aeruginosa are a consistent and rich source of anti-phage defense systems. *Nucleic acids research* (2023). https://doi.org:10.1093/nar/gkad317

Acknowledgements
S.S. was supported by the Damon Runyon Fellowship Award (DRG 2352-19). Experiments were funded by awards to J.B.D by the Vallee, Searle, and Kleberg Foundations. K.S.M. and E.V.K. are supported by intramural funds of the US Department of Health and Human Services (National Institutes of Health, National Library of Medicine). We thank Jason Peters for technical discussions, Andy Millard for the Enterobacteriophage genome dataset, and Ry Young, Andrew Fire, Carol Gross and Jonathan Weissman for guidance.

Author Contributions
Conceptualization, S.S., J.B.D.; Methodology, S.S., E.L.; Resources, A.F., D.S.; Software, S.S.; Investigation, S.S., H.C., D.S.G.; Formal Analysis, S.S., H.C., M.J., K.S.M., E.V.K.; Visualization, S.S., H.C., M.J., K.S.M.; Writing – Original Draft, S.S., H.C.; Writing – Reviewing & Editing, S.S., H.C., M.B., E.V.K., J.B.D.; Supervision, S.S., J.B.D.; Funding Acquisition, J.B.D.

Competing Interests
S.S. is co-founder and equity holder in BillionToOne, Inc. J.B.D. is a scientific advisory board member of SNIPR Biome, Excision Biotherapeutics, and LeapFrog Bio, consults for BiomX, and is a scientific advisory board member and co-founder of Acrigen Biosciences. The Bondy-Denomy lab received research support from Felix Biotechnology.

Additional Information
Supplementary Information is available for this paper. Mass Spectrometry data available at PRIDE https://www.ebi.ac.uk/pride/login with Username: reviewer_pxd038604@ebi.ac.uk and Password: 83OmeYpD. High-throughput sequencing data available at SRA with accession PRJNA952709. Correspondence and requests for materials should be addressed to Sukrit.Silas@ucsf.edu, Joseph.Bondy-Denomy@ucsf.edu. Reprints and permissions information is available at www.nature.com/reprints.

Supplementary Figure 1. Conditional-lethal phenotypes of known R-M inhibitors in wild *E. coli.* (A) Schematic representation of conditional-lethality data from Figure 1d, showing only known R-M inhibitors and an uncharacterized AG, *orf7*:*aop1*. **(B)** AGs from (A) tested for their ability to block EcoKI restriction of phage λvir in plaque assays with restriction-competent *E. coli* MG1655. **(C)** Various mutants of Ocr tested in MG1655 for their ability to block EcoKI restriction of λvir, and in ECOR22 and ECOR55 for their ability to trigger PCD activity of EcoR22I and Ronin respectively. F54V mutant of Ocr (known to evade a previously described ARIA, PARIS) still triggers ARIA activity in ECOR55 but not ECOR22. Numbered Ocr mutants were described elsewhere[19]. Notably, Ral, which operates by allosterically hyperactivating the methyltransferase[61] (unlike Ocr and ArdA) does not trigger PCD in any strains. Conversely, Aop1 elicits PCD from the same systems triggered by Ocr and ArdA in (A) but does not inhibit Type I R-M activity in (B). The predicted structure of Aop1 does not resemble a DNA mimic (Supplementary File 7), and the protein is not negatively charged (pI ~4 for Ocr/ArdA, but ~8 for *aop1*), suggesting a distinct mechanism of R-M inhibition. Thus, depending on their mechanism, some R-M inhibitors trigger PCD in these strains, whereas others do not.

Supplementary Figure 2. Transposon insertions that suppress AG-induced conditional-lethal phenotype in ECOR22. Distribution of unique transposon insertions recovered from ECOR22 survivors from Tn5 libraries upon *ocr* expression, mapped to the EcoR22I locus.

Supplementary Figure 3. Upstream non-coding region is required for Ronin PCD activity. (A) The first 135-bp of the 235-bp upstream non-coding region (marked by a thick red line) are required for activity as truncation down to 100-bp causes inactivation. ArdA is chromosomally integrated under pLlacO-1 control and strains in the left (ArdA repressed) and right (ArdA induced) panels are isogenic. **(B)** Over-expression of 235-bp upstream region is not toxic in TOP10 indicating that it is not the PCD effector. **(C)** As in (A), PCD phenotype of truncated Ronin system when overexpressed (indicated as "pBAD"). Over-expression of truncated Ronin did not rescue the PCD defect indicating that the upstream region contributes something in addition to the promoter.

Supplementary Figure 4. Phage-host combinations included in AG screens (A) Phage susceptibility of select wild ECOR strains. Numbers in boxes are log10-transformed infection scores (7 = 10^7-fold higher plaquing than 0). Borders indicate that individual plaques were visible in that phage-host combination. Combinations with attenuated phage infection that were selected for the AG screen are in yellow. Combinations with attenuated infection that were attempted but yielded no survivors in liquid infection experiments (perhaps due to Abi phenotypes) are in red. **(B)** Phage-host combinations where counter-defense phenotypes were confirmed for any AGs. Red denotes combinations where the screen could not be completed (as in (A)). Yellow: combinations that were tested where none of the 200 AGs produced counter-defense effects. Green: combinations where phage-sensitizing effects could not be verified by plaque assays. Blue: combinations where counter-defense phenotypes were confirmed by plaque assay (at least 100X higher EOP upon AG expression). **(C)** Defense repertoire of selected ECOR strains. Computational prediction of the presence or absence of various defense systems in ECOR strains used in the AG screen. Predictions were generated using the ISLAND software suite[62].

Supplementary Figure 5. Over-expression of *galU* reverses T4 sensitization by *gnarl3*. log10-transformed T4 infection scores in wild hosts ECOR1 (green), ECOR3 (orange), ECOR22 (blue) with or without UDP-glucose biosynthesis pathway genes *galU, galE, galF,* and ECA precursor *wecB* cloned onto a plasmid and overexpressed.

Supplementary File 1. Worked example of AG-finding algorithm. Readme.txt describes how to run the code. Sample database of Salmonella phage genomes is included. All ARs from this phage dataset are reported in the output file SPFMphages_accRegions_nr_all.txt. Any AR can be extracted using the supplied code and then visualized using Clinker[60]. The AR shown in Figure 1d is extracted by default.

Supplementary File 2. Manually inspected accessory regions from Enterobacteriophage genomes. Visualizations of all ~200 accessory regions (ARs) that were manually inspected. The first slide in the file shows the layout for the rest, with the AR score calculated according to the formula in Methods, followed by a unique identifier for each hypothetical phage group ("phamily"). Each AR was also given a short, memorable name (top right). Slides with green backgrounds indicate ARs from which genes were selected for the AG screen. Authors' impressions upon AR inspection are included in the Notes section.

Supplementary File 3. Phage accessory genes selected for AG screens. 200 accessory genes (AGs) selected for screening. Genomic locations, protein and phage genome identifiers, and protein descriptions are indicated. IDs denoted by "UDP" correspond to the orf/AG numbers used throughout the manuscript.

Supplementary File 4. Experimental record of AG screens. A detailed record of all AG screens. The first sheet contains phage-host combinations, dates of the experiment, and preliminary results. The next three sheets each correspond to one replicate each of AG screening. Empirically measured multiplicity at time of infection, and dilutions used before plating are also recorded. Screens that had to be repeated for technical reasons are color-coded.

Supplementary File 5. Hits from transposon-mediated PCD suppressor screens. Follow-up screens to find inactivating transposon insertions in putative Abi mechanisms that might trigger PCD in response to conditional-lethal AGs. These screens are performed with RB-TnSeq libraries prepared with ECOR hosts already containing IPTG-inducible trigger-AGs (no phages present). Genes inactivated by transposon insertions that were overrepresented in the AG-on condition (IPTG-induced) relative to the AG-off condition (Glucose-repressed) are reported as hits.

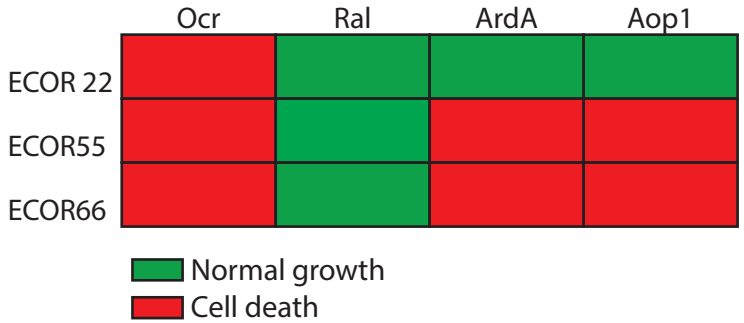
Supplementary File 6. Genomic architectures of all ECOR55-like Ronin systems. Spreadsheet contains accession numbers and genomic locations of all *ronA* homologs. Grouping variables indicate whether a Type I R-M system (containing *hsdR, hsdM,* and *hsdS*) was found anywhere in the genome, and also whether a full (>400 aa) DUF4102 integrase was present immediately upstream of Ronin. HTML files split Ronin systems into two groups based on the presence of the integrase, and gene annotations can be viewed by mouse-over.

Supplementary File 7. AlphaFold predictions for all AGs that produced phenotypes in AG screens. PDB files containing predicted structures of *orf7* (*aop1*; triggers PCD by Ronin), *orf24* (triggers PCD in ECOR61), *orf48*, *orf63*, *orf92*, (*gnarl1-3*; putative O-antigen modifiers), *orf98* (triggers PCD in ECOR22), *orf116* (*abc1*; triggers PCD by P4 prophage), *orf126* (broad-spectrum counter-defense), *orf143* (IpII; inhibitor of GmrSD), *orf145* (triggers PCD in MG1655), *orf148* (broad-spectrum counter-defense), and *orf184* (triggers PCD by gop-beta-cII).

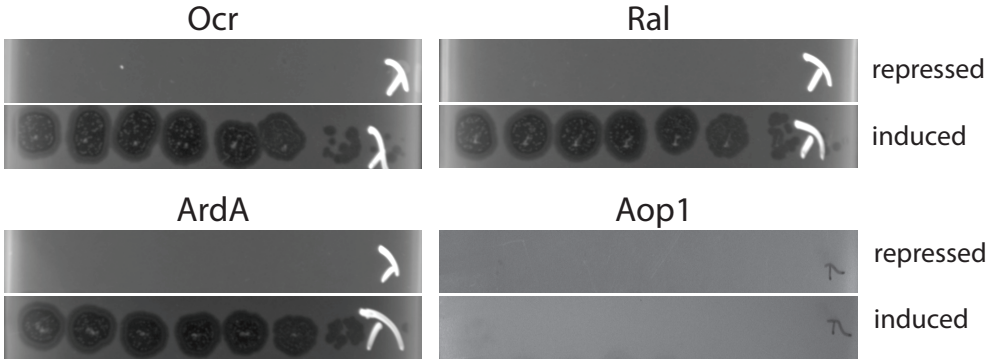
Supplementary File 8. Experimental record of follow-up transposon whole-genome knockout screens. The first sheet contains a detailed record of all follow-up transposon screens designed to phenocopy counter-defense phenotypes of AGs. These screens are

1382 performed with RB-TnSeq libraries prepared with naïve ECOR hosts (no AGs present). Each
1383 library is challenged by various phages one at a time. The timeline of each experiment indicates
1384 library growth characteristics. Empirically measured multiplicity at time of infection, and dilutions
1385 used before plating are also recorded. Screens were performed in two stages: first, a pilot with
1386 one phage for each RB-TnSeq library, and then a larger experiment with all other phages that
1387 showed enhanced infection upon expression of a counter-defense AG. The second sheet lists
1388 all hits recovered from the screens.
1389
1390 **Supplementary File 9. Host-AG combinations for affinity-purification/mass-spectrometry.**
1391 Matrix of all host-AG combinations where potential binding partners of AGs were affinity-purified
1392 from the native host expressing counter-defense AGs, and identified by mass-spectrometry.
1393 Experiments were performed in triplicate. *Orf74* produced no phenotypes during AG screening
1394 and was used as a control for all experiments in the relevant hosts.
1395
1396 **Supplementary File 10. Accessory regions of AGs selected for further study.**
1397 Visualizations of ARs which contain AGs that produced counter-defense or conditional-lethal
1398 phenotypes in AG screens.

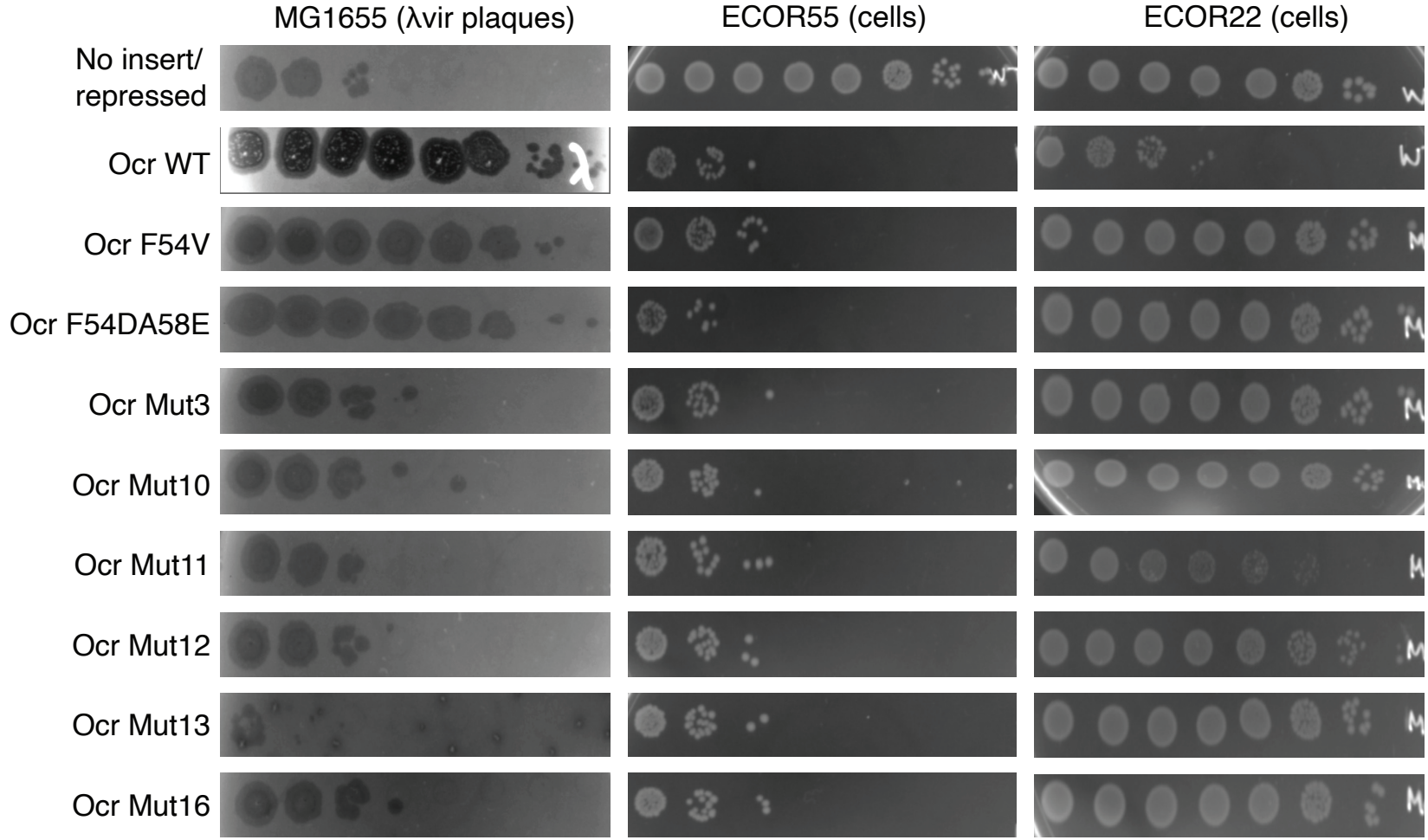
A. R-M inhibitors trigger abortive defenses in wild E. coli

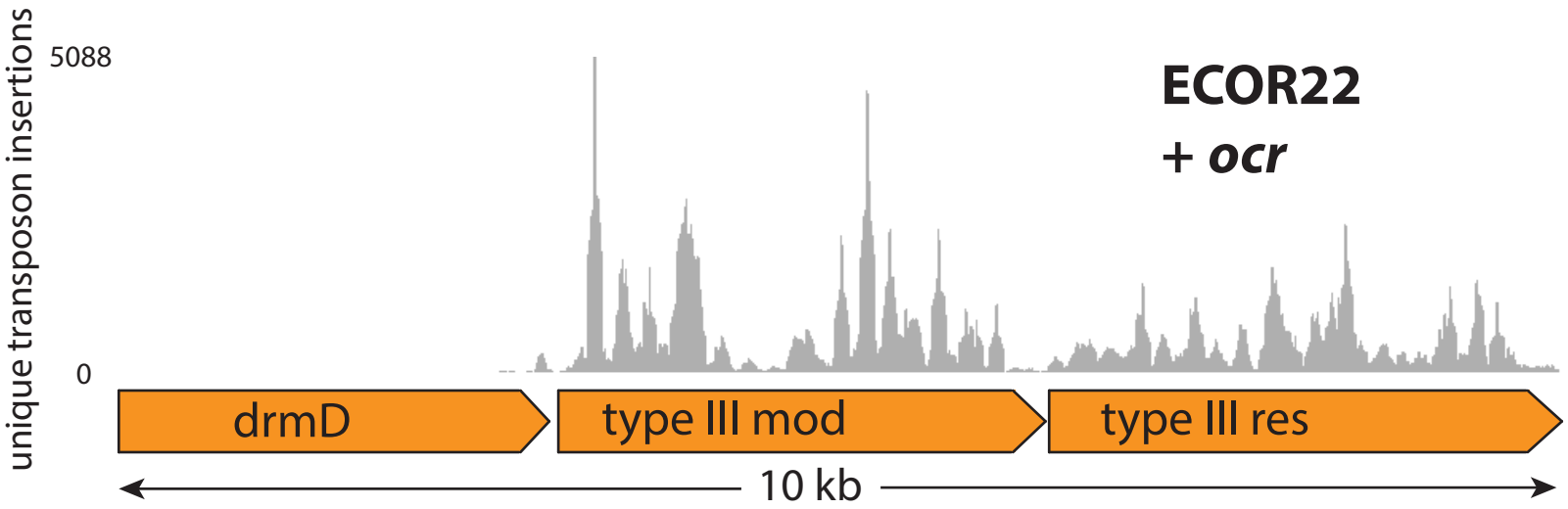


B. R-M inhibition by AGs in MG1655

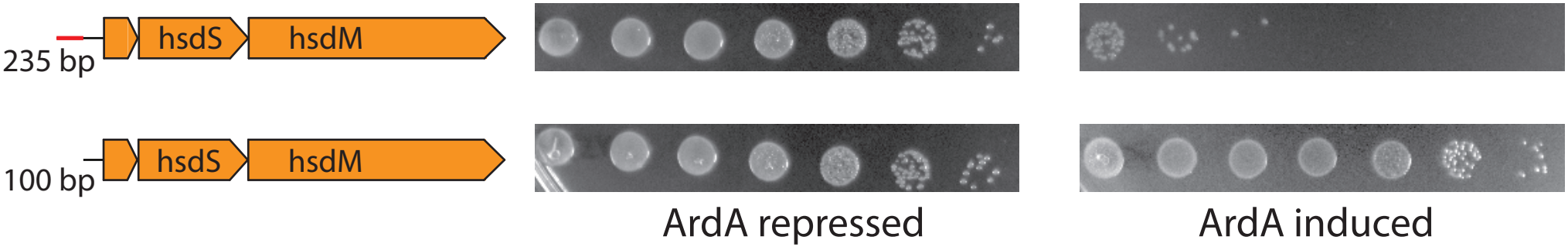


C. Effect of Ocr mutations on R-M (plaquing on MG1655) and PCD activity (ECOR55, 22)

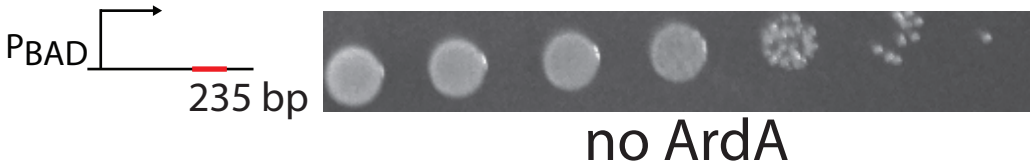



unique transposon insertions
5088
0
ECOR22
+ ocr
drmD
type III mod
type III res
10 kb

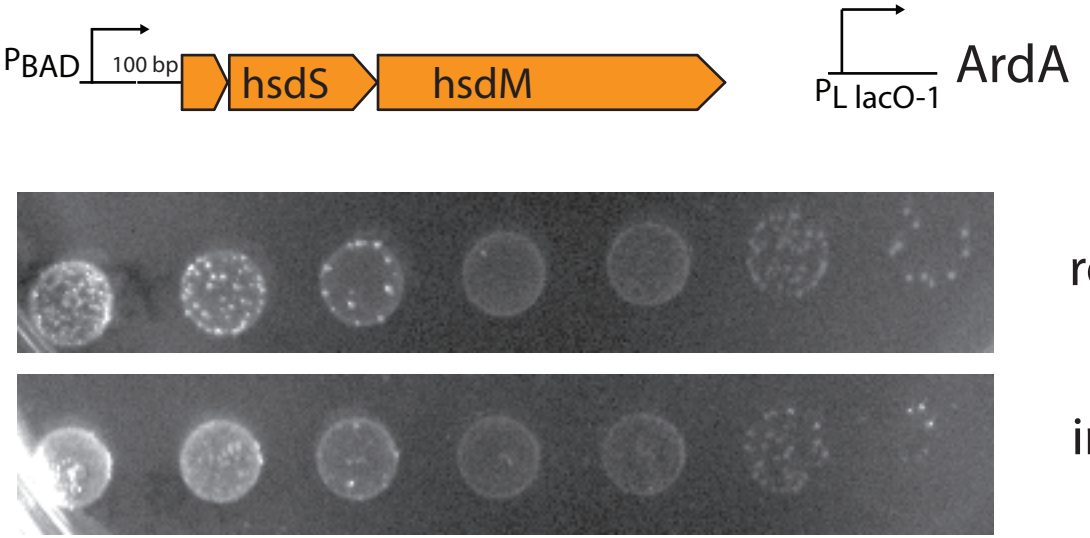
A. Ronin upstream region is required for PCD activity



B. Ronin upstream region is not toxic when overexpressed on its own



C. Over-expression of Ronin with truncated upstream region



A. Phage susceptibility of ECOR strains

	T2	T3	T4	T5	T6	λvir	P1vir	T7
C990	6	6	3	7	6	7	6	7
MG1655	6	4	6	7	6	3	6	7
ECOR1							3	
ECOR3	2				3		2	
ECOR4	4	4		7	6			7
ECOR6								
ECOR10	0		0		0			
ECOR13	6	2	6	4	7	7	3	7
ECOR15	0				1			
ECOR18					1		1	
ECOR19					1		1	
ECOR21	1		2				1	
ECOR22	0		3		0		2	
ECOR24					0		3	
ECOR25								
ECOR51			4		6		2	
ECOR53			1		2		2	
ECOR55			1		5		2	
ECOR59			0		1		1	
ECOR61	2	0	1	3	1		2	0
ECOR66			3		5		3	

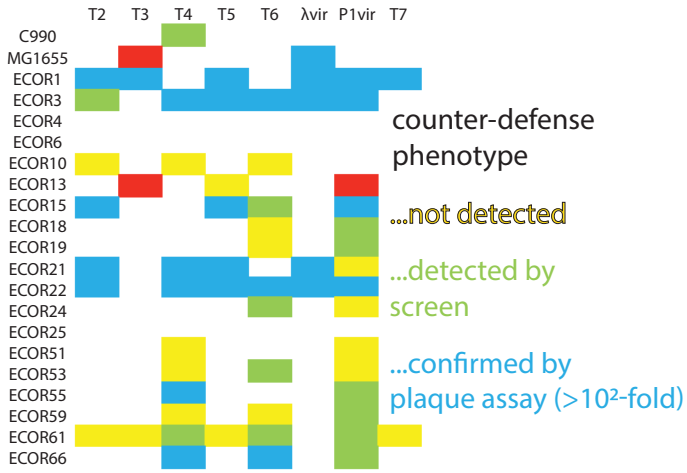
n EOP = n

n EOP = n, and single plaques visible

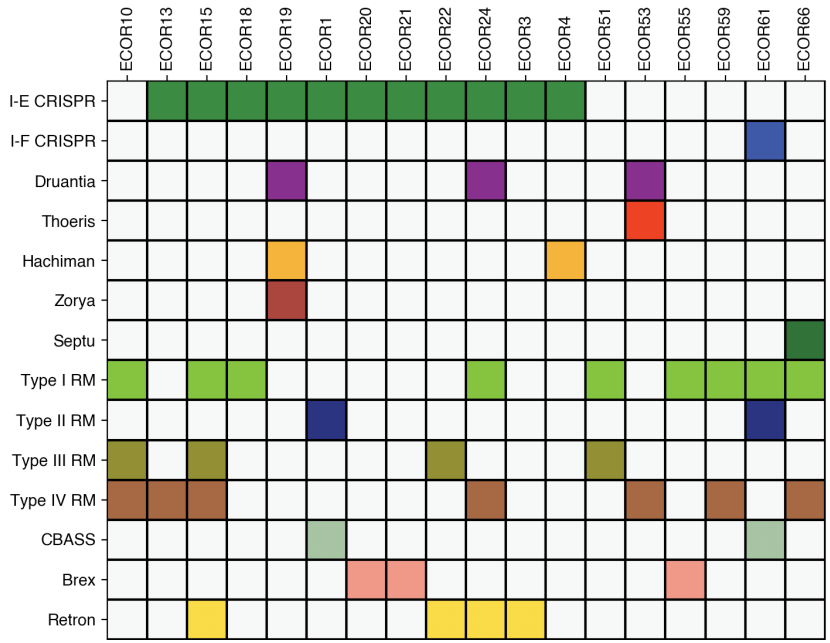
selected for PhAGENT screen

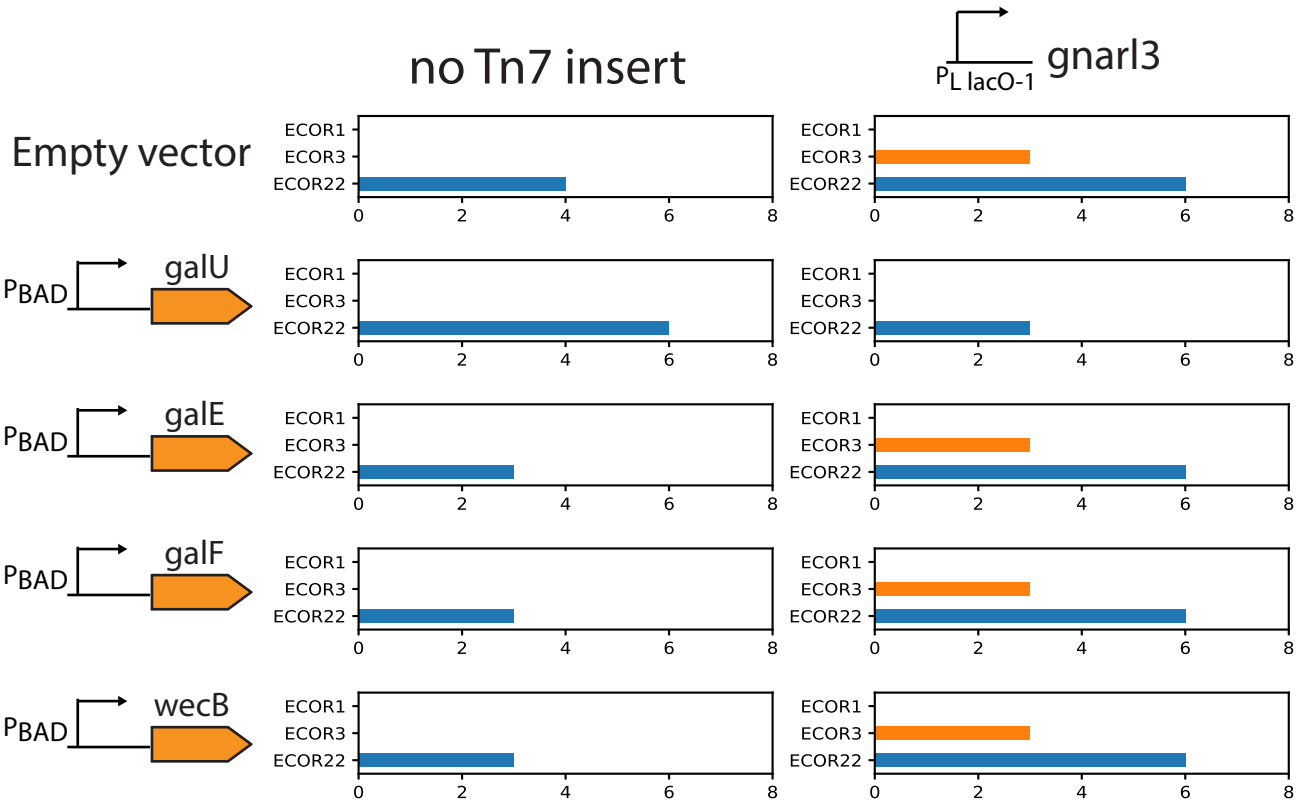
attempted but screen could not be performed

B. Phage-host combinations with confirmed counter-defense phenotypes



C. Defense systems in ECOR strains




no Tn7 insert
PL lacO-1
gnarl3
Empty vector
PBAD
galU
galE
galF
wecB
ECOR1
ECOR3
ECOR22
0
2
4
6
8