Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages

Fatima Aysha Hussain, Javier Dubert, Joseph Elsherbini, Mikayla Murphy, David Vaninsberghe, Philip Arevalo, Kathryn Kauffman, Bruno Kotska Rodino-Janeiro, Hannah Gavin, Annika Gomez, et al.

Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages

Authors: Fatima Aysha Hussain[1]†, Javier Dubert[1,2]†, Joseph Elsherbini[1], Mikayla Murphy[1], David VanInsberghe[1], Philip Arevalo[1], Kathryn Kauffman[1], Bruno Kotska Rodino-Janeiro[1,3], Hannah Gavin[1], Annika Gomez[1], Anna Lopatina[3], Frédérique Le Roux[4,5]*, and Martin F. Polz[1,3]*.

Affiliations:

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology; 77 Massachusetts Avenue, Cambridge, MA, USA.

[2]Department of Microbiology and Parasitology, University of Santiago de Compostela; Spain.

[3]Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, Center for Microbiology and Environmental Systems Science, University of Vienna; Austria.

[4]Ifremer, Unité Physiologie Fonctionnelle des Organismes Marins, ZI de la Pointe du Diable, CS 10070, F-29280 Plouzané, France

[5]Sorbonne Universités, UPMC Paris 06, CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688 Roscoff cedex, France

*Corresponding authors. Emails: martin.f.polz@univie.ac.at and fleroux@sb-roscoff.fr

† These authors contributed equally to this work.

Abstract:

Although it is generally accepted that phages drive bacterial evolution, how these dynamics play out in the wild remains poorly understood. Here we show that susceptibility to viral killing in marine *Vibrio* is mediated by large and highly diverse mobile genetic elements. These phage defense elements display exceedingly fast evolutionary turnover, resulting in differential phage susceptibility among clonal bacterial strains while phage receptors remain invariant. Protection is cumulative, and a single bacterial genome can harbor 6-12 elements, accounting for over 90% of the flexible genome amongst close relatives. The rapid turnover of these elements decouples phage resistance from other genomic features. Thus, resistance to phages in the wild follows alternative evolutionary trajectories than those predicted from lab-based evolutionary experiments.

Main Text:

Bacterial viruses (phages) are ubiquitous across Earth's biosphere and control microbial populations through predatory interactions (*1–3*). Because successful killing depends on molecular interaction with the host, phages display higher specificity than most other microbial predators. This is an important reason for renewed interest in the clinical use of phages as alternatives to antibiotics (*4*). Like antibiotics, however, phage killing exerts strong selection for resistance in bacterial hosts (*5*). Consequently, how bacteria naturally acquire resistance has important implications for understanding microbial community dynamics as well as the long-term success of phage therapy. Laboratory co-evolution studies have consistently identified phage-receptor mutations as key drivers of resistance (*6, 7*), with genetic analyses indicating secondary contributions by restriction-modification (RM) (*8–10*) and abortive infection (Abi) systems (*11–13*). However, because phages target important surface structures, such as the lipopolysaccharide (LPS) and membrane proteins as receptors, it is questionable whether mutations in receptors represent primary adaptive strategies in complex microbial communities since such modifications frequently incur significant fitness costs (*6*). Indeed, many additional defense mechanisms have recently been discovered, including clustered regularly interspaced short palindromic repeats (CRISPR) systems (*14*), and several other innate immunity mechanisms, many of which remain to be mechanistically characterized (*15–19*). Importantly, genes encoding both receptors and defense systems have been shown to occur frequently in variable genomic islands (*6, 20–22*), or associated with mobile genetic elements (*23–26*). However, while for CRISPR-based adaptive immunity, which is present in over 40% of bacteria (*27*), evolutionary processes have been documented on ecological timescales (*28, 29*), much less is known for innate immunity mechanisms, which defend most bacteria in the wild (*30*). Neither

the predominant mechanisms of resistance to phages nor the dynamics of resistance gain and loss are well understood (*31, 32*), limiting our knowledge of how phage predation structures diversity and drives evolution of wild microbial populations.

Here we combine population genomic and molecular genetic approaches to determine
60   how phage resistance arises in natural microbial populations lacking adaptive immunity. We recently created a large genomically-resolved phage-host cross-infection network, using sympatric environmental isolates (*33–35*). The Nahant Collection was established in the context of a 93-consecutive-day coastal-ocean time series (*36*) and comprises over 1,300 strains of marine *Vibrio* ranging in relatedness from near-clonal to species-level divergence (*34, 37*).
65   *Vibrio* hosts isolated on three different days (ordinal dates 222, 261, 286) were used as "bait" to isolate 248 co-occurring lytic viruses by quantitative plaque assays, and each of 245 plaque positive hosts were subsequently challenged with all phage isolates to establish an all-by-all cross-infection matrix (Fig. S1A). Phages tend to be highly specific, as indicated by the general sparsity of the matrix. Additionally, even the most closely related hosts, differentiated by a few
70   single nucleotide polymorphisms (SNPs) across the core genome, are preyed upon by different phages (Fig. S1B). Furthermore, this extends to broader host-range phages that, although capable of infecting multiple hosts, are typically limited to a single strain within a host clade (Fig. S1C).

Because the observation that nearly clonal bacterial strains are subject to differential predation suggests extremely rapid evolution of phage resistance, we sought to identify the
75   mechanisms in an exemplary set of 22, nearly clonal isolates of *Vibrio lentus* (Fig. 1A, left) that are differentially infected by two highly divergent groups of lytic siphovirus phages consisting of four and 18 isolates, respectively ("orange" and "purple" in Fig. 1A, top; Fig. S2A-D, Data S1). Testing all pairwise interactions of these phages with all potential hosts at varying multiplicity of

infection (MOI), reveals two phenotypes: (i) host lysis and phage production and (ii) host lysis only at high phage concentrations without phage production (Fig. S3). The latter effect, termed "lysis from without" **(38)**, along with observed adsorption of representative phages to differing hosts (Fig. S4), suggests that the phages can adsorb to a broader set of hosts than they can kill.

Supporting our hypothesis, we find that while the two phage types use different receptors, all bacterial genomes encode both sets of receptors. Molecular genetic analysis suggests the "orange" phage receptor is the Type II secretion system (T2SS) pseudopilus (GspH), while the "purple" phages have two receptors, one being LPS, an extracellular receptor, and the second being a sodium transporter (NqrC), a membrane receptor (**(39)**, Fig. S5, Tables S1 and S2). Comparative genomics showed that every identified gene encoding a phage receptor is identical across all 23 "orange", "purple" and outgroup strains. The noteworthy exception is the single outgroup strain (10N26154A12) that differs by one SNP in *gspN*: N201K, which prevents a single "orange" phage 1.127.O from attaching (Fig. 1A). Although this suggests receptor mutations play a secondary role, the data overall support that receptors are not primarily responsible for the observed phage specificity, leading us to explore potential intracellular mechanisms of host resistance by a combination of comparative genomics and molecular genetics.

Clustering the pangenome of the 22 near clonal *Vibrio lentus* isolates using high-quality genomes, together with a slightly more divergent outgroup for additional genomic context, (Fig. 1B, right) reveals the presence-absence patterns of large flexible genomic regions containing putative phage defense genes. Using a k-mer-based approach to conduct all-by-all pair-wise genome comparisons (Data S2), we identified a total of 28 flexible genome regions (>5kb in length), 26 of which show clear hallmarks of being mobile genetic elements (MGEs) and

harboring at least one gene related to a known phage defense gene (Data S3). These elements range in size from 8,500 to 89,600 bp and contain genes annotated as recombinases/integrases and transposases (Data S3), but the elements appear overall unrelated since only few genes are shared (Fig. 1C, Data S3). Except for two episomal elements, one of which is a prophage, all others are chromosomally integrated and their presence-absence patterns in different clonal genomes indicate that these regions are transferred as independent units (Fig. S6). Of the chromosomally integrated elements, three are also putative prophages. We refer to these 26 MGEs as phage defense elements (PDEs) since no other function obviously relevant to the host cell can be identified. Their number ranges between six and 12 in each strain, and collectively, the PDEs account for over 95% of the flexible genomic regions (Fig. 1D, Data S3). Even if only known defense genes, excluding the entire PDEs are considered, phage defense genes account for >20% of the flexible gene content (Fig. 1D). We observe a similar range of 12-17% in diverse *Vibrio* species in our collection (Fig. S7) even though these comprise higher genomic diversity than the clonal strains and have known flexible genome elements other than those involved in phage defense (*40*). If we consider that most of the remaining PDE genes may be involved in maintenance and spread, and that we did not detect any other functional annotations on the PDEs, we predict a large, if not the largest, portion of the flexible genome might be involved in phage defense.

To investigate the impact of PDEs on phage defense, we looked more closely at a near clonal subset of these hosts that have differential presence of PDEs matched by phage predation profiles. These hosts consist of a group of four strains and another group of 15 strains that are only differentiated by 14 SNPs across the core genome, yet are differentially infected by the two phage groups (Fig. 1A, Fig. S8, Table S3). The flexible genome of these "orange" and "purple"

hosts is differentiated by three (PDEs 1-3) and two (PDEs 4-5) large PDEs (Fig. 1E). PDE1, PDE4, and PDE5 have Type 1 RM systems, and PDE2 and PDE3 have several putative defense genes. These include many genes related to abortive infection (Abi) systems, as well as genes with domains homologous to retrons, reverse transcriptases, and GTPases (*17, 22*). A striking feature of the PDEs is that their insertion does not appear to disrupt host functions. For example, PDE1 inserts into "orange" hosts' 5'-deoxynucleotidase nucleosidase (*yfbr*) gene, thereby truncating it, but encodes its own divergent copy of the same gene. Similarly, although PDE2 disrupts a gene encoding a thiol peroxidase upon insertion, it encodes a copy in the middle of the element (Fig. 1E). These five PDEs are also found in more divergent *Vibrio* genomes, which indicates they transfer via horizontal gene transfer (Fig. S9), sometimes inserting into different genomic neighborhoods, although at conserved insertion motifs with the exception of a few SNPs (Fig. S10).

We asked to what extent do each of the PDEs contribute to the differential resistance observed between the "purple" and "orange" hosts? All three PDE-encoded RM systems (PDEs1,4 and 5; Fig. 1E) appear active based on methylome sequencing analysis (Table S4, **(*39*)**). To further characterize the contribution to resistance of other potential defense genes, we focused on the three PDEs specific to "orange" host strains as an exemplar. First, we found that all three PDEs are transcribed during an infection with a "purple" phage (Fig. S11). Second, we systematically knocked out large regions of the three PDEs containing putative defense as well as unidentified candidate genes hypothesized to drive resistance of a representative "orange" host strain (10N26155C8) to "purple" phages (Fig. 2A), and then challenged the knockouts with all "purple" and "orange" phages (Fig. 2B, Fig. S12; Fig. S13). This genetic analysis shows that in addition to the RM-encoding PDE, both PDEs encoding alternative defense systems are also

active and together contribute to phage resistance since deletion of all three PDEs is required for full sensitivity.

150   The complexity of the interaction among all PDEs leading to full resistance is illustrated by deletion experiments covering all possible combinations. Knocking out RM-containing PDE1 alone increased the "lysis from without" phenotype tenfold (from 10^{-1} to 10^{-2} phage dilution), but did not yield viable phage progeny when assayed using a re-streaking test (Fig. 2C). This phenotype led us to hypothesize that there may be a multi-level defense structure involving the

155   remaining two PDEs. Knocking out PDE2 and PDE3 independently resulted in no change from the wild type (WT) phenotype. However, knocking out PDE2 and PDE3 together allowed for both killing and propagation of the phage at a wider range of high MOIs ($\leq 10^{-5}$ phage dilution). This phenotype is consistent with restriction modification escape mutants (**(39)**, Fig. S14). Knocking out PDE1 and PDE2 together resulted in the same tenfold increase in killing observed

160   when deleting PDE1 alone, while knocking out PDE1 and PDE3 together yields cell death at much lower MOIs (10^{-5} phage dilution), but importantly, in neither case were viable phages propagated (Fig. 2C). These data suggest that both PDEs afford protection but that PDE3 provides stronger resistance than PDE2. Finally, knocking out all three PDEs simultaneously resulted in the "orange" strain becoming just as susceptible to the "purple" phages as are

165   "purple" WT host strains. Therefore, we conclude all three elements are needed for full, WT-level defense.

Further genetic characterization of defense genes within the PDEs underscores the diversity of undiscovered or poorly characterized genes with anti-phage function. We took advantage of our double PDE mutants to scrutinize the remaining active PDE for phage defense.

170   While we were only able to find a single gene (the restriction enzyme) necessary for defense on

PDE1 (Figs. 2A, S15), three unannotated genes and two putative GTPases within a module on the 5'end of PDE2 appear involved in defense (Figs. 2A, S16). For PDE3, we inactivated 11 different genes as well as two regions associated the two retrons present; however, none showed a loss of resistance phenotype. We therefore analyzed the host transcriptome during infection with a "purple" phage and identified 14 highly expressed and/or co-transcribed regions (Fig. S17A). Using the triple knockout mutant, ΔΔΔPDE123, we complemented eight of these regions with their native promoter and observed restoration of phage defense by region R3 containing a gene with DUF2971 (Figs. 2A, S17B,C). Hence two of the three PDEs carry uncharacterized phage defense genes.

That receptor genes are invariant across the two near clonal host groups challenges the notion that receptor variation is primarily responsible for resistance (*6*, *20*). Thus, we asked to what extent these receptors are variable across more diverse populations. Populations are defined here as gene flow clusters that also represent ecological units (*37*, *41*). The recognition of population boundaries is key for interpretation of gene or allele frequency in the context of differential selective forces. Surprisingly, looking across 107 *Vibrio* isolates, spanning 10 populations, all putative receptors are highly monomorphic at the population level indicating they are under purifying selection. The two genes identified as receptors in *V. lentus* are identical, or nearly so, at the amino acid sequence level across populations and thus well below the average diversity of core genes (Fig. S18A). This is corroborated by phylogenetic trees based on nucleotide sequences showing that all members of each population carry the same or highly similar gene variants, a pattern consistent with the occurrence of recent gene-specific selective sweeps. A notable exception is *gspH* which is more diverse in two of the populations (Fig. S18B). Finally, the LPS, which frequently serves as the primary receptor for many phages, also

appears similar at the population level since the genes responsible for its synthesis display

population-specific presence/absence patterns suggesting the synthesis pathway is conserved

(Fig. S18C). Thus, although putative receptors can reside in variable regions when more

divergent genomes are compared (*20*), their evolution appears to be constrained when population

structure is considered. This constraint may arise because in wild populations, these surface

structures are optimized for ecological interactions, and indicates a key difference from the lab

where, in rich media, receptor mutations frequently arise in phage-host co-cultures (*42*). This

observed invariance suggests that other selective forces that compete with phage resistance play

an important role in receptor evolution in the wild. This interpretation is consistent with

predictions that intracellular defenses should be important under such conditions (*43*, *44*), as well

as with phage-host evolution in nature differing from the lab (*45*). It is possible that receptor-

mediated defenses may only be advantageous under extreme predation regimes, or under regimes

of low effective diversity, such as a clonal infection or laboratory culture.

Defense being largely determined by differential PDE carriage at the clonal level

suggests that the rate of turnover is surprisingly fast and confirms theoretical considerations that

resistance genes should be mobile because the cost of resistance limits their utility under

changing predation pressures (*25*). Genomic comparisons of the 23 strains illustrates this high

turnover since genomes with zero SNPs can differ by at least one putative PDE, and the "purple"

and "orange" hosts groups carry a minimum difference of five PDEs while being separated by

only 14 SNPs (Fig. 1 and Fig. S8). Further, across the time series of a mere 90 days, we observe

a differential shift in relative PDE abundance. We designed droplet digital PCR assays to span

the insertion sites of three of the PDEs to constrain the detection of the resistance elements to

closely related genetic backgrounds (Fig. S18). While PDE3 and PDE4, representing "orange"

and "purple" genetic backgrounds, respectively, are near perfectly correlated across the entire time series, PDE2 only gains in importance towards the end of the time series. These dynamics suggest rapid evolution at the population level, either from acquisition of PDE2 or a shift in abundance of host clones with differential PDE carriage. Both scenarios reflect a rapidly shifting selective landscape of host-phage interactions.

PDEs structuring host defense among near clonal isolates is likely general beyond the detailed description provided here. Within other *Vibrio* species in our collection, clonal isolates also differ in their phage predation profiles (Fig. S1), and, even among bacterial pathogens that – in contrast to vibrios – follow a primarily non-recombinogenic mode of evolution (*37, 46–48*), we observe similarly rapid turnover of putative phage defense elements, several of which are prophages themselves. When analyzing publicly available genomes with identical ribosomal protein gene sequences from *Listeria* (Fig. S20, Data S4), *Salmonella* (Fig. S21, Data S5)*, and *Clostridium* (Fig. S22, Data S6), we observe a surprisingly similar dominance (~90%) of the flexible genome associated with phage defense regions across these diverse species.

We have shown that innate immunity mechanisms can be rapidly turned over on the population level, with three major implications. First, because co-existing diversity in many natural bacterial populations is high (*49*), resistance at the clonal level effectively lowers prey concentration and thus phage-host encounter. Second, encoding resistance on highly mobile elements can make phage resistance a fast-changing phenotypic trait independent of physiological or metabolic traits encoded by the core genome. This allows the core genome to be maintained over the long-term even in the face of phage predation. In other words, while phages may exert negative frequency dependent selection on their hosts, this selection is largely acting at the level of MGEs. Finally, rapid acquisition of resistance may also have to be considered for

repeated or longer-term use of phages in therapy since resistance may be easily acquired and quickly spread through bacterial populations, just as the connection to MGEs (i.e., primarily plasmids) has led to an unanticipated rise in antibiotic resistance. An important open question is therefore what the mechanisms of transfer and life history strategies of PDEs are. Together, our findings suggest that phage resistance is an important, if not the most important, selective force determining clonal bacterial diversity, with phage defense elements potentially explaining a very large portion of the previously enigmatic bacterial flexible genome.

References and Notes:

1. K. E. Wommack, R. R. Colwell, Virioplankton: Viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000).

2. M. Breitbart, F. Rohwer, Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* **13**, 278–284 (2005).

3. H. G. Hampton, B. N. J. Watson, P. C. Fineran, The arms race between bacteria and their phage foes. *Nature*. **577**, 327–336 (2020).

4. K. E. Kortright, B. K. Chan, J. L. Koff, P. E. Turner, Phage Therapy: A renewed approach to combat antibiotic-resistant bacteria. *Cell Host Microbe*. **25**, 219–232 (2019).

5. S. J. Labrie, J. E. Samson, S. Moineau, Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317–327 (2010).

6. S. Avrani, O. Wurtzel, I. Sharon, R. Sorek, D. Lindell, Genomic island variability facilitates Prochlorococcus-virus coexistence. *Nature*. **474**, 604–608 (2011).

7. J. R. Meyer, D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick, R. E. Lenski, Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science*. **335**, 428–432 (2012).

8. W. Arber, Host-controlled modification of nacteriophage. *Annu. Rev. Microbiol.* **19**, 365–378 (1965).

9. G. G. Wilson, N. E. Murray, Restriction and modification systems. *Annu. Rev. Genet.* **25**, 585–627 (1991).

10. G. Bertani, J. J. Weigle, Host controlled variation in bacterial viruses. *J. Bacteriol.* **65**, 113–121 (1953).

11. I. J. Molineux, Host-parasite interactions: recent developments in the genetics of abortive phage infections. *New Biol.* **3**, 230–236 (1991).

12. L. Snyder, Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol. Microbiol.* **15**, 415–420 (1995).

270   13.   A. Lopatina, N. Tal, R. Sorek, Abortive Infection: Bacterial suicide as an antiviral immune strategy. *Annu. Rev. Virol.* **7**, 371–384 (2020).

14.   R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, P. Horvath, CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. **315**, 1709–1712 (2007).

275   15.   S. Doron, S. Melamed, G. Ofir, A. Leavitt, A. Lopatina, M. Keren, G. Amitai, R. Sorek, Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*, **359**, 6379 (2018).

16.   A. Millman, A. Bernheim, A. Stokar-Avihail, T. Fedorenko, M. Voichek, A. Leavitt, Y. Oppenheimer-Shaanan, R. Sorek, Bacterial retrons function in anti-phage defense. *Cell*. **183**, 1551-1561 (2020).

17.   L. Gao, H. Altae-Tran, F. Böhning, K. S. Makarova, M. Segel, J. L. Schmid-Burgk, J. Koob, Y. I. Wolf, E. V. Koonin, F. Zhang, Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science*. **369**,
280   1077–1084 (2020).

18.   R. P. Novick, G. E. Christie, J. R. Penadés, The phage-related chromosomal islands of Gram-positive bacteria. *Nat. Rev. Microbiol.* **8**, 541–551 (2010).

19.   A. A. Govande, B. Duncan-Lowey, J. B. Eaglesham, A. T. Whiteley, P. J. Kranzusch, Molecular basis of CD-NTase nucleotide selection in CBASS anti-phage defense. *Cell Rep.* **35**, 109206 (2021).

285   20.   F. Rodriguez-Valera, A.B. Martin-Cuadrado, B. Rodriguez-Brito, L. Pasic, T. F. Thingstad, F. Rohwer, A. Mira, Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* **7**, 828–836 (2009).

21.   K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* **193**, 6039–6056 (2011).

290   22.   A. Bernheim, R. Sorek, The pan-immune system of bacteria: antiviral defence as a community resource. *Nat. Rev. Microbiol.*, 1–7 (2019).

23.   N. D. McDonald, A. Regmi, D. P. Morreale, J. D. Borowski, E. F. Boyd, CRISPR-Cas systems are present predominantly on mobile genetic elements in Vibrio species. *BMC Genomics*. **20**, 105 (2019).

24.   A. C. McKitterick, K. D. Seed, Anti-phage islands force their target phage to directly mediate island excision
295   and spread. *Nat. Commun.* **9**, 2348 (2018).

25.   E. V. Koonin, K. S. Makarova, Y. I. Wolf, M. Krupovic, Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. *Nat. Rev. Genet.* **21**, 119–131 (2020).

26.   K. N. LeGault, S. G. Hays, A. Angermeyer, A. C. McKitterick, F. Johura, M. Sultana, T. Ahmed, M. Alam, K. D. Seed, Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts. *Science*. **373**,
300   6554 (2021).

27.   K. S. Makarova, D. H. Haft, R. Barrangou, S. J. J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. M. Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost, E. V. Koonin, Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477 (2011).

28.   A. F. Andersson, J. F. Banfield, Virus population dynamics and acquired virus resistance in natural microbial
305   communities. *Science*. **320**, 1047–1050 (2008).

29.   E. Laanto, V. Hoikkala, J. Ravantti, L.-R. Sundberg, Long-term genomic coevolution of host-parasite interaction in the natural environment. *Nat. Commun.* **8**, 111 (2017).

30. T. Dimitriu, M. D. Szczelkun, E. R. Westra, Evolutionary ecology and interplay of prokaryotic innate and adaptive immune systems. *Curr. Biol. CB*. **30**, 1189–1202 (2020).

31. A. Chevallereau, B. J. Pons, S. van Houte, E. R. Westra, Interactions between bacterial and phage communities in natural environments. *Nat. Rev. Microbiol.*, 1–14 (2021).

32. S. van Houte, A. Buckling, E. R. Westra, Evolutionary ecology of prokaryotic immune mechanisms. *Microbiol. Mol. Biol. Rev.* **80**, 745–763 (2016).

33. K. M. Kauffman, J. M. Brown, R. S. Sharma, D. VanInsberghe, J. Elsherbini, M. Polz, L. Kelly, Viruses of the Nahant sollection, characterization of 251 marine *Vibrionaceae* viruses. *Sci. Data*. **5**, 180114 (2018).

34. K. M. Kauffman, F. A. Hussain, J. Yang, P. Arevalo, J. M. Brown, W. K. Chang, D. VanInsberghe, J. Elsherbini, R. S. Sharma, M. B. Cutler, L. Kelly, M. F. Polz, A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature*. **554**, 118–122 (2018).

35. K. M. Kauffman, W. K. Chang, J. M. Brown, F. A. Hussain, J. Y. Yang, M. F. Polz, L. Kelly, Resolving the structure of phage-bacteria interactions in the context of natural diversity. *BioRxiv* p. 2021.06.27.449121, (2021), doi:10.1101/2021.06.27.449121.

36. A. M. Martin-Platero, B. Cleary, K. Kauffman, S. P. Preheim, D. J. McGillicuddy, E. J. Alm, M. F. Polz, High resolution time series reveals cohesive but short-lived communities in coastal plankton. *Nat. Commun.* **9**, 266 (2018).

37. P. Arevalo, D. VanInsberghe, J. Elsherbini, J. Gore, M. F. Polz, A reverse ecology approach based on a biological definition of microbial populations. *Cell*. **178**, 820-834 (2019).

38. M. Delbrück, The growth of bacteriophage and lysis of the host. *J. Gen. Physiol.* **23**, 643–660 (1940).

39. Supplementary results, materials, and methods are available as supplementary materials on Science Online.

40. O. X. Cordero, M. F. Polz, Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* **12**, 263–273 (2014).

41. B. J. Shapiro, J. Friedman, O. X. Cordero, S. P. Preheim, S. C. Timberlake, G. Szabó, M. F. Polz, E. J. Alm, Population genomics of early events in the ecological differentiation of bacteria. *Science*. **336**, 48–51 (2012).

42. E. R. Westra, S. van Houte, S. Oyesiku-Blakemore, B. Makin, J. M. Broniewski, A. Best, J. Bondy-Denomy, A. Davidson, M. Boots, A. Buckling, Parasite exposure drives selective evolution of constitutive versus inducible defense. *Curr. Biol.* **25**, 1043–1049 (2015).

43. S. Zborowsky, D. Lindell, Resistance in marine cyanobacteria differs against specialist and generalist cyanophages. *Proc. Natl. Acad. Sci.*, U.S.A. 116, 16899–16908 (2019).

44. E. O. Alseth, E. Pursey, A. M. Luján, I. McLeod, C. Rollie, E. R. Westra, Bacterial biodiversity drives the evolution of CRISPR-based phage resistance. *Nature*. **574**, 549–552 (2019).

45. C. A. Hernandez, B. Koskella, Phage resistance evolution in vitro is not reflective of in vivo outcome in a plant-bacteria-phage system. *Evolution*. **73**, 2461–2475 (2019).

46. K. E. Dingle, D. Griffiths, X. Didelot, J. Evans, A. Vaughan, M. Kachrimanidou, N. Stoesser, K. A. Jolley, T. Golubchik, R. M. Harding, T. E. Peto, W. Fawley, A. S. Walker, M. Wilcox, D. W. Crook, Clinical *Clostridium difficile*: Clonality and pathogenicity locus diversity. *PLoS One*. **6**, 19993 (2011).

47. B. J. Shapiro, How clonal are bacteria over time? *Curr. Opin. Microbiol.* **31**, 116–123 (2016).

48. Y. Yin, W. Tan, G. Wang, S. Kong, X. Zhou, D. Zhao, Y. Jia, Z. Pan, X. Jiao, Geographical and longitudinal analysis of *Listeria* monocytogenes genetic diversity reveals its correlation with virulence and unique evolution. *Microbiol. Res.* **175**, 84–92 (2015).

49. J. R. Thompson, S. Pacocha, C. Pharino, V. Klepac-Ceraj, D. E. Hunt, J. Benoit, R. Sarma-Rupavtarm, D. L. Distel, M. F. Polz, Genotypic Diversity Within a Natural Coastal Bacterioplankton Population. *Science.* **307**, 1311–1313 (2005).

50. Code is provided in Data S2 and online (DOI: 10.5281/zenodo.5228182).

51. Supporting data for phylogenetic trees provided in Data S9 and online (DOI: 10.5061/dryad.sj3tx965k).

52. D. E. Hunt, L. A. David, D. Gevers, S. P. Preheim, E. J. Alm, M. F. Polz, Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science.* **320**, 1081–1085 (2008).

53. S. G. John, C. B. Mendez, L. Deng, B. Poulos, A. K. M. Kauffman, S. Kern, J. Brum, M. F. Polz, E. A. Boyle, M. B. Sullivan, A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ. Microbiol. Rep.* **3**, 195–202 (2011).

54. K. M. Kauffman, M. F. Polz, Streamlining standard bacteriophage methods for higher throughput. *MethodsX.* **5**, 159–172 (2018).

55. N.F. Alikhan, N. K. Petty, N. L. Ben Zakour, S. A. Beatson, BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics.* **12**, 402 (2011).

56. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

57. J. Huerta-Cepas, K. Forslund, L. P. Coelho, D. Szklarczyk, L. J. Jensen, C. von Mering, P. Bork, Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).

58. J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernández-Plaza, S. K. Forslund, H. Cook, D. R. Mende, I. Letunic, T. Rattei, L. J. Jensen, C. von Mering, P. Bork, eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, 309–314 (2019).

59. L. Guy, J. R. Kultima, S. G. E. Andersson, genoPlotR: comparative gene and genome visualization in R. *Bioinformatics.* **26**, 2334–2335 (2010).

60. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).

61. J. P. Meier-Kolthoff, M. Göker, VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinforma. Oxf. Engl.* **33**, 3396–3404 (2017).

62. P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, S. Hunter, InterProScan 5: genome-scale protein function classification. *Bioinformatics.* **30**, 1236–1240 (2014).

63. S. R. Eddy, Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, 1002195 (2011), doi:10.1371/journal.pcbi.1002195.

64. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics.* **10**, 421 (2009).

385   65. M. Baym, S. Kryazhimskiy, T. D. Lieberman, H. Chung, M. M. Desai, R. Kishony, Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One*. **10**, 0128036 (2015).

66. https://github.com/rrwick/Filtlong.

67. M. Kolmogorov, J. Yuan, Y. Lin, P. A. Pevzner, Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).

390   68. R. R. Wick, M. B. Schultz, J. Zobel, K. E. Holt, Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. **31**, 3350–3352 (2015).

69. http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

70. R. R. Wick, L. M. Judd, C. L. Gorrie, K. E. Holt, Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13**, 1005595 (2017).

395   71. D. Hyatt, G.L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, L. J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. **11**, 119 (2010).

72. T. J. Treangen, B. D. Ondov, S. Koren, A. M. Phillippy, The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **15**, 524 (2014).

73. A. K. M. Kauffman, thesis, Massachusetts Institute of Technology (2014).
400   https://dspace.mit.edu/handle/1721.1/90046.

74. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

75. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).

405   76. S. V. Angiuoli, S. L. Salzberg, Mugsy: fast multiple alignment of closely related whole genomes. *Bioinforma. Oxf. Engl.* **27**, 334–342 (2011).

77. R Core Team, R: A language and environment for statistical computing (2019). https://www.R-project.org.

78. G. Marçais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. **27**, 764–770 (2011).

410   79. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).

80. B. D. Ondov, T. J. Treangen, P. Melsted, A. B. Mallonee, N. H. Bergman, S. Koren, A. M. Phillippy, Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).

81. M. Bastian, S. Heymann, M. Jacomy, paper presented at the Third International AAAI Conference on Weblogs and Social Media, San Jose, CA, 17-20 May 2009.

415   82. L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).

83. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402
420   (1997).

84. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. **28**, 3150–3152 (2012).

85. I. A. Murray, T. A. Clark, R. D. Morgan, M. Boitano, B. P. Anton, K. Luong, A. Fomenkov, S. W. Turner, J. Korlach, R. J. Roberts, The methylomes of six bacteria. *Nucleic Acids Res*. **40**,11450–11462 (2012).

86. R. J. Roberts, T. Vincze, J. Posfai, D. Macelis, REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res*. **43**, D298-299 (2015).

87. L. Ferrières, G. Hémery, T. Nham, A.-M. Guérout, D. Mazel, C. Beloin, J.-M. Ghigo, Silent mischief: Bacteriophage mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J. Bacteriol.* **192**, 6418–6427 (2010).

88. S. L. Chiang, E. J. Rubin, Construction of a mariner-based transposon for epitope-tagging and genomic targeting. *Gene.* **296**, 179–185 (2002).

89. F. Le Roux, J. Binesse, D. Saulnier, D. Mazel, Construction of a *Vibrio splendidus* mutant lacking the metalloprotease gene *vsm* by use of a novel counter-selectable suicide vector. *Appl. Environ. Microbiol.* **73**, 777–784 (2007).

90. S. Das, J. C. Noe, S. Paik, T. Kitten, An improved arbitrary primed PCR method for rapid characterization of transposon insertion sites. *J. Microbiol. Methods*. **63**, 89–94 (2005).

91. F. M. Lauro, K. Tran, A. Vezzi, N. Vitulo, G. Valle, D. H. Bartlett, Large-scale transposon mutagenesis of *Photobacterium profundum* SS9 reveals new genetic loci important for growth at low temperature and high pressure. *J. Bacteriol.* **190**, 1699–1709 (2008).

92. Y. Jiao, A. Kappler, L. R. Croal, D. K. Newman, Isolation and characterization of a genetically tractable photoautotrophic Fe(II)-oxidizing bacterium, *Rhodopseudomonas palustris* Strain TIE-1. *Appl. Environ. Microbiol.* **71**, 4487–4496 (2005).

93. M.-E. Val, O. Skovgaard, M. Ducos-Galand, M. J. Bland, D. Mazel, Genome engineering in *Vibrio cholerae*: A feasible approach to address biological issues. *PLoS Genet.* **8**, 1002472 (2012).

94. A. Lemire, D. Goudenège, T. Versigny, B. Petton, A. Calteau, Y. Labreuche, F. Le Roux, Populations, not clones, are the unit of vibrio pathogenesis in naturally infected oysters. *ISME J.* **9**, 1523–1531 (2015).

95. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

96. H. Wildschutte, S. P. Preheim, Y. Hernandez, M. F. Polz, O-antigen diversity and lateral transfer of the w*be* region among *Vibrio splendidus* isolates. *Environ. Microbiol.* **12**, 2977–2987 (2010).

97. J. Söding, A. Biegert, A. N. Lupas, The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).

480   **List of Supplementary materials:**

Materials and Methods

Figures S1 to S22

Tables S1 to S4

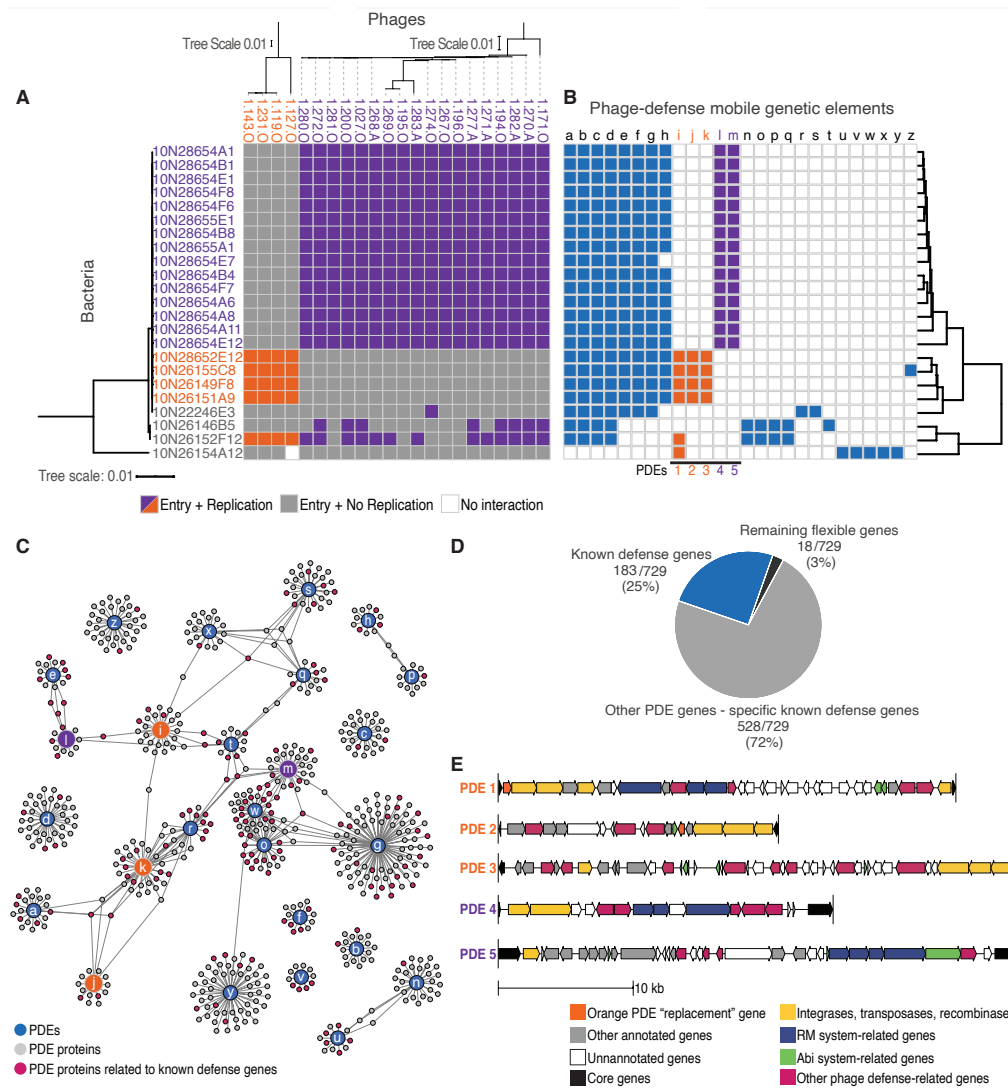References (52-97)

485   Data S1 to S9

490

Fig. 1. Near-clonal strains of *Vibrio lentus* differ in sensitivity to phage predation and in their carriage of mobile genetic elements encoding for phage defense genes. (A) Phage host-range matrix with rows representing bacterial strains and columns representing phages. (A-left) Core genome phylogenetic tree of hosts, (A-top) whole genome tree of viruses. (B) Distribution of 26 PDEs in 23 hosts with PDEs responsible for defense of the "orange" and "purple" host groups indicated in the same colors, (B-right) hierarchical clustering of presence/absence of genomic regions in bacterial hosts corroborating "orange" and "purple" phage infection patterns. (C) Shared genes (>50% amino acid sequence identity) among the 26 PDEs displayed as a network where gray nodes are genes with defense genes indicated in pink. (D) Fraction of the bacterial flexible genome attributed to phage defense. Amongst the 23 clones, an all-by-all genomic comparison shows 97% of flexible regions greater than 5kbp are putative PDEs. Only 25% of the total flexible genes match known defense genes while the majority are other PDE-associated genes (72%), many of which are unannotated. The remaining genes, not associated with PDEs, account for 3% of the flexible genome. (E) Gene diagrams of mobile genetic elements specific to the "orange" and "purple" host groups.
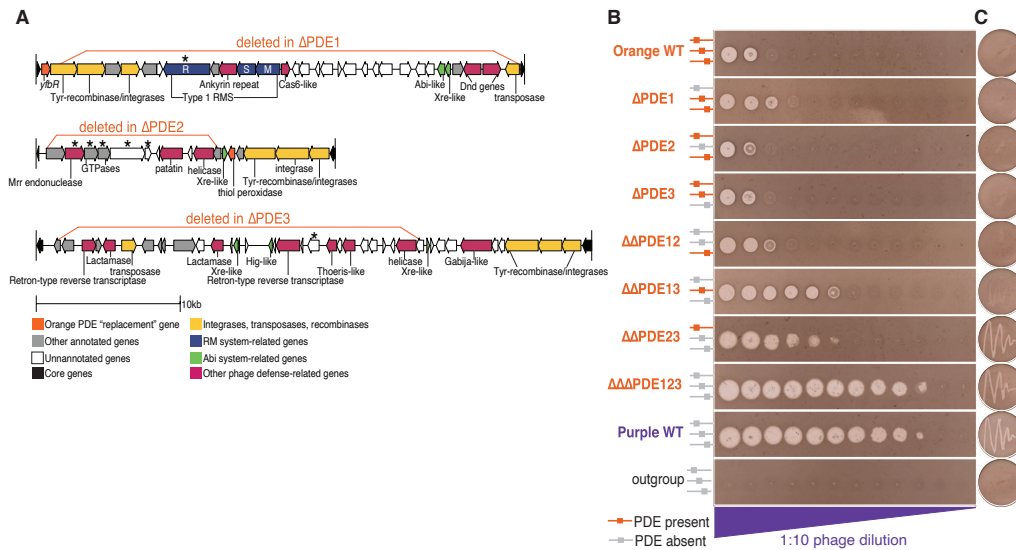
Fig. 2. Changes in susceptibility to phage killing observed for phage defense element (PDE) deletions. (A) Detailed gene diagrams indicate marker-less deletions of defense-gene containing portions of the different "orange" PDEs as well as genes determined to be responsible for defense phenotypes (marked with *, Figs. S15-17). (B) Lawns of bacterial hosts with drop spots of a 1:10 dilution series of "purple" phage (image shows representative "purple" phage 1.281.O; plating completed in triplicate for each phage, Fig. S12). Cartoons on the left depict presence or absence of different PDEs in each strain. From top to bottom: "orange" wild type host (10N26155C8), ΔPDE1, ΔPDE2, ΔPDE3, ΔΔPDE12, ΔΔPDE13, ΔΔPDE23, ΔΔΔPDE123, "purple" wild type host (10N28654F7, positive control), outgroup (10N26149C11, negative control). (C) Re-streak test for propagation of phage progeny from drop spot clearings. Only infections of ΔΔPDE23, ΔΔΔPDE123, and "purple" wild type hosts produce viable phages, indicated by clearing on the re-streak plates.

Supplementary Materials for

Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages

Fatima Aysha Hussain, Javier Dubert, Joseph Elsherbini, Mikayla Murphy, David VanInsberghe, Philip Arevalo, Kathryn Kauffman, Bruno Kotska Rodino-Janeiro, Hannah Gavin, Annika Gomez, Anna Lopatina, Frédérique Le Roux*, and Martin F. Polz*
*Correspondence to: martin.f.polz@univie.ac.at and fleroux@sb-roscoff.fr

Materials and Methods

Bacteria and phage isolation

Bacteria and phages were obtained in a previous study from coastal seawater collected from Canoe Cove, Nahant, MA, USA, on August 22 (ordinal day 222), September 18 (261), and October 13 (286), 2010 (*34*). *Vibrio* bacteria were isolated using a size fractionation approach, followed by plating on selective media, as described previously (*52*). Briefly, to capture bacteria associated with large particles and zoo- and phytoplankton, seawater was filtered through a 63 µm average pore size plankton net. To capture bacteria in smaller size fractions, including small particles and smaller zoo- and phytoplankton as well as bacteria occurring in the free-living fraction, water pre-filtered through a 63 µm net was serially passed through 5 µm, 1 µm, and 0.2 µm polycarbonate filters. To isolate vibrios from each fraction, material captured in the plankton net and on filters was resuspended in artificial seawater (ASW; Sea Salts from Sigma-Aldrich), and the suspensions passed through polyether sulfone 0.2 µm filters. These final filters were placed directly on agar plates of MTCBS (Difco Thiosulfate-Citrate-Bile-Sucrose Agar amended with 10 g/L of NaCl to final concentration of 2% w/v) to allow for selective growth of *Vibrio* colonies. Colonies were purified by serial passaging on agar plates of: first, TSB2 (Tryptic Soy Broth, 1.5% Difco Bacto Agar, amended with 15 g NaCl to 2% w/v); second, MTCBS, and third TSB2. Colonies were inoculated into 1 mL of Difco 2216 Marine Broth (2216MB) in 96-well 2 mL culture blocks and allowed to grow, shaking at room temperature, for 48 hours. Glycerol stocks for preservation at -80°C were prepared by combining 100 µL of culture with 100 µL of 50% glycerol (50% water) in 96-well microtiter plates. The naming of each strain reflects isolation location, day, and size fraction: 10N refers to the 2010 collection of samples from Nahant; 222, 261, 286 are the ordinal dates of the year; 54, 55, 56 are three replicates of the 63 µm fraction; 51, 52, 53, are three replicates of the 5 µm fraction; 48, 49, 50 are three replicates of the 1 µm fraction; and 45, 46, 47 are three replicates of the 0.2 µm or free-living fraction. The final portion of the name is the original storage well in a 96-well plate. The "orange" phage isolates were, with one exception, collected on day 261 and distributed between the 1 µm, 5 µm, and 63 µm fraction, while all "purple" isolates were collected on day 286 from the 63 µm fraction. This isolation scheme reflects a snapshot of the lifestyle of the isolated bacteria, where 5 µm and 63 µm represent living in a particle-attached state in the ocean.

For phage isolation, 4 L of seawater was collected in triplicate on each day in the time series and separately filtered through a Sterivex 0.22 µm barrel filter into a sterile 4 L collection bottle using a peristaltic pump. Phages were directly concentrated from this filtrate using an iron flocculation and filtering method described previously (*53*). Briefly, iron (III) chloride, which is spiked into the sample, precipitates phages from the solution, and then the precipitates are

collected onto 90 mm 0.2 µm polycarbonate filters using a glass cup-frit system. Precipitates are finally dissolved in 4 mL of oxalate solution to yield a quantitative concentration of 1,000x from the original 4 L. The final phage concentrate was stored at 4ºC in the dark until used to isolate specific viruses for different bacterial hosts.

Vibrio isolates were used as "bait" to obtain phages from the concentrates using direct plating in soft agar overlays. Plaques from the bait assay were archived frozen in 2216MB and glycerol and phages for use in the host range assay were subsequently randomly selected from archives for each host and purified by triple serial passage using tube-free agar overlays (*33*, *54*) on their hosts of isolation. Phages were amplified on their hosts of isolation using primary small-scale liquid cultures inoculated with plaque plugs from their final serial passage in agar overlays, followed by plating of primary lysates into agar overlays to achieve high-titer stocks. Top agar layers of lawns "at confluence" (saturated with plaques but not completely cleared) were harvested into 2216MB, centrifuged at 5,000 x *g* for 20 minutes, and filtered through Sterivex 0.22 µm barrel filters to generate the lysates used for the all-by-all host range cross test as well as for phage DNA extraction and sequencing (*33*), as well as methylation profiling (described below).

Phage characterization
Circular representations of the previously sequenced phage genomes (Fig. S2CD) were generated using BRIG (*55*) with publicly available NCBI GenBank files; annotations were made based on manual review of GenBank predictions and supplemented with Phyre2 (*56*) and EggNog-Mapper (*57*, *58*) annotation. Genome diagrams (Fig. S2EF) were generated using the GenoPlotR package in R (*59*) with predicted protein coding genes indicated as arrows colored to correspond to protein sequence clusters, as defined using default settings of MMseqs2 (*60*); and with "orange" (Fig. S2E) and "purple" (Fig. S2F) phages clustered and identified as two separate genus-level groups using the D6 amino acid OPTSIL clustering algorithm in the VICTOR classifier (*61*) with whole genome concatenated protein sequences.

Although siphoviruses are frequently temperate, there is no evidence the "orange" and "purple" phages integrate into bacterial genomes, nor of genetic features associated with a temperate life history strategy (Data S1). All phages in the Nahant Collection were evaluated for lysogeny associated gene features and presence as prophages in bacterial genome sequences previously (*35*) and verified here. No such evidence was found for any of the "purple" or "orange" phages. Evaluation for lysogeny associated genes: EggNOG-mapper v.2 (*58*) and InterProScan v.5.39-77.0 (*62*) annotations for all phages were reviewed for integrases (PF00589, PF00239), replicative transposases (PF02914), and transcriptional repressor domains (IPR010982, IPR010744, IPR001387, IPR032499); iterative sequence-based searches using the jackhmmer function of HMMER v.3.1.2 (*63*) were used to screen for plasmid associated genes SopA (NP_046923.1) and ParA (AAA99230.1). No such genes were found for any "orange" or "purple" phages. Evaluation for integration as prophages: BLASTn (*64*) searches were used to screen 309 Nahant Collection bacterial genome assemblies for sequences with significant similarity to those of the "orange" and "purple" phages. No significant hits (e-value ≤1.00E-05) were found between any of the phages and the bacterial genomes (Data S1).

Hybrid assemblies of bacterial genomes
Because we noticed that PDEs assemble poorly when genomes were sequenced with short read technology, we used Illumina short reads and Pacific Biosciences (PacBio) long reads in

combination to assemble high quality (nearly closed) genomes. For short read sequencing, bacterial isolates were grown overnight from a single colony in 1.2 mL of 2216MB in deep-well blocks and processed in bulk. Genome libraries were prepared for sequencing using the Nextera DNA Library Preparation Kit (Illumina) with 1-2 ng input DNA per isolate, as previously described (*65*). Genomes were sequenced on 100 bp pared-end sequencing runs using Illumina HiSeq, with 50-60 samples multiplexed per lane. When available, Illumina HiSeq short reads from previous work (*34*) were used, otherwise new short read data were generated for this study. High quality bacterial genomic DNA for PacBio sequencing was prepared separately. A single colony was inoculated into a 250 mL Erlenmeyer flask with 50 mL of 2216MB and grown shaking at room temperature for 24 hours. The fresh culture was pelleted by centrifugation at 5,000 x *g* for 20 minutes and then immediately processed for DNA extraction when possible, or frozen at -20°C for short term storage. The Qiagen Genomic Tip 500x kit was used following the manufactures guidelines, and the final DNA was collected by spooling using a glass rod rather than centrifugation to avoid shearing. DNA was stored in 500 µL of elution buffer at 4°C for 24-48 hours to allow for full resuspension before sequencing at either the Yale Center for Genome Analysis (PacBio RS II, without multiplexing) or the BioMicroCenter at MIT (Sequel, with multiplexing).

A custom hybrid assembly pipeline was designed to process the data. Briefly, PacBio reads were filtered at different length cutoffs using Filtlong (*66*) and then assembled using Flye (*67*) to create a set of reference genomes. The reference genomes were visualized using Bandage (*68*) and the best genome was selected, based on completion and coverage, to be used as a reference in the final assembly. For the final assembly, Illumina reads were trimmed using Trim-Galore (*69*), PacBio reads were quality filtered with the trimmed Illumina reads using Filtlong, and both sets of processed reads, together with the best Flye assembly, were used as inputs for the Unicycler (*70*) assembler.

Bacterial genome annotations
Genomes were annotated using Prodigal 2.6 (*71*) for Open Reading Frame (ORF) prediction. Predicted ORFs were annotated using InterProScan5 (*62*) using the iprlookup, goterms, and pathways options. InterProScan5 matches against 13 databases by default, which are listed here: https://github.com/ebi-pf-team/interproscan/wiki/HowToRun#included-analyses. Two optional databases were included for this analysis: TMHMM for predicted transmembrane proteins and SignalP for predicted signal peptide cleavage sites.

Host relationships
Phylogenetic relationships among the *Vibrio* strains were estimated by comparison of (i) concatenated alignments of ribosomal proteins and *hsp60* as described below in methods for, "Phage host-range matrix" (Fig. S1A), and (ii) core genome alignments. For estimation of whole genome relationships, we used the Parsnp program (*72*) with the recombination flag (-x) to construct whole genome SNP trees. Then, HarvestTools was used to convert from a ggr format to a snp fasta file, and finally, IQ-Tree was used to optimize the final tree. This was done for the 23 clones (Fig. 1A, left) as well as for the "purple" and "orange" strains alone (Fig. S8). SNPs in the core genome were located using custom python scripts and verified by visualizing on Ginger, the Harvest graphic user interface. SNP details for Fig. S8 are given in Table S3.

Phage host-range matrix

Initial phage host-range was determined in a previous study (*34, 35*). Briefly, all *Vibrio* strains for which at least one phage was found ("plaque positive") were used in the host-range assay and challenged with all phages purified as described above. Bacterial hosts were plated in agar overlays in large 150 mm plates and stamped with phage lysates arranged in triplicate in 96-well arrays using 96-spot blotters (BelArt, Bel-blotter 96-tip replicator, 378760002). Clearing seen in at least 2/3 replicates was scored as a positive kill (*73*). The concentration of each lysate in the original assay was not normalized to allow for higher throughput, however, the assay was repeated for select hosts of interest at a range of concentrations (see methods for "varying phage concentrations" below).

To organize bacterial hosts in the matrix by phylogeny, concatenation of ribosomal proteins and *hsp60* sequences was used to construct a phylogenetic tree reflecting the relationship of the core genome (Fig. S1A). When genome sequences were available, we used HMMER (*63*) to find ribosomal proteins, and aligned the sequences with MAFFT (*74*). Amino acid sequences of *hsp60* proteins were also extracted from genomes via HMMER using pfam PF00118. The *hsp60* sequences were aligned using the mafft-fftnsi algorithm. When genomes were not available, *hsp60* sequences that were Sanger-sequenced were added to this alignment using the mafft-fftnsi algorithm with the –addfragments option. The *hsp60* alignment and the ribosomal protein alignment were concatenated and used to create the phylogenetic tree in Figure S1A using RAxML (options: –q, -m GTRGAMMAX) (Data S9) (*75*).

Host range assays at varying phage concentrations
To determine the host-ranges of the specific phages used in this work at a higher resolution, we re-assayed a subset of hosts and phages of interest using a range of concentrations. Bacterial hosts were grown in 5 mL of 2216MB overnight from single colonies streaked on 1.5% Bacto Agar plates supplemented with 2216MB. Phage lysates were prepared as described above and diluted in 2216MB to form a ten-fold dilution series from 10^0-10^{-7}. Five µL drop spots of each dilution were pipetted onto bacterial host lawns made using a tube-free agar overlay method (*54*) and incubated at room temperature for 24 hours before evaluating phage entry and efficiency of plating at the varying concentrations (Fig.1A). Plates with different killing were imaged using a flatbed scanner (Epson Perfection V800 Photo Scanner - Product No. B11B22320) and captured using VueScan Software by Hamrick (Fig. S3). Lysis from without was only observed when using this higher resolution assay, thus the authors recommend performing such an assay when evaluating phage host-range whenever possible.

Phage adsorption assay
To determine if all phages were attaching to all hosts, for "orange" phage 1.143.O and "purple" phage 1.281.O, we compared the number of free phages remaining in solution after exposure to "orange" host 10N26155C8, "purple" host 10N28654F7, an unrelated *Vibrio* (outgroup) control 10N26149C11, and a no-host negative control (2216MB). Three different colonies of each bacterial strain were each inoculated in 3 mL of 2216MB and grown shaking at 25°C for 4 hours. Bacterial concentration was estimated at optical density measured at 600 nm wavelength (OD_{600}), and each replicate was normalized to OD_{600} of 0.3 followed by 100-fold dilution. One mL of each diluted culture was aliquoted into individual wells of a 96-well culture block and bacteria were grown shaking at room temperature for another 3.5 hours to reach mid-exponential phase. Twenty µL of phage lysate was added to each well at varying concentrations (ranging from MOI = 0.001 to 10 phages/bacteria on average) and staggered in time to achieve an

adsorption time of 30 minutes (Fig. S4). After allowing phages to adsorb, 200 µL of the phage and bacteria mixture was filtered using a 96-well filter system (Millipore MultiScreen Vacuum Manifold) to remove bacteria and any infecting or adsorbed phages. Five µL of a ten-fold dilution series of each well was then drop spotted onto a fresh lawn of a sensitive host (orange strain 10N26155C8 was used for experiments with 1.143.O and 10N28654F7 was used for experiments with 1.281.O) made in rectangular petri dishes (1-well Nunc Rectangular Dishes, Polystyrene, Sterile by Thermo Scientific (Supplier No. 267060)). Plates were incubated at room temperature for 18-24 hours and then imaged using a flatbed scanner as described above (Fig. S4). Phage adsorption was estimated by comparing the number of plaque forming units (PFUs) in each dilution series to the no host control. For example, in Figure S4A the same order of magnitude of PFUs is evident in both the outgroup and the no-host control, meaning there is no phage adsorption for the outgroup. Yet, there are an order of magnitude more PFUs in both controls compared to the "purple" and "orange" hosts, implying equal adsorption is seen on the "purple" and "orange" hosts.

Identification, annotation, and characterization of putative PDEs in the flexible genome

In order to determine the differences in the flexible genome amongst the 23 clonal strains (Fig. 1), we created a multiple alignment using Mugsy (*76*), and performed a hierarchical clustering of the alignment blocks, greater than 500 bp, by length in R using hclust (*77*) (Fig.1B, right).

Then to specifically identify the exact regions of interest in the flexible genome, we used a k-mer based approach (Data S2), followed by careful manual curation. All pairs of genomes were compared. First, each genome was split into 31-mers using Jellyfish (*78*), then shared kmers between the genomes being compared were removed and only unique kmers were mapped back to the reference genomes they originated from using Bowtie2 (*79*). Any unique region greater than 1,000 bp was kept and a gap of 3,000 bp was allotted to account for genes that may have been shared between the two genomes splitting a complete region. Regions were checked for duplication and the largest region of any overlapping regions was saved. We then used Mash (*80*) to compare all the unique regions to each other and clustered any region greater than 5kbp with minimum Jaccard similarity of 0.95. We visualized the clustering using Gephi (*81*) and then chose one representative from each cluster by hand to make a final list of unique regions. We then used BLASTn (*64*) and custom python scripts to determine which genomes harbored which elements with >95% identity and >80% length considered a match. We removed any element that appeared in all genomes. Then we went back through the elements and identified insertions sites (see below) to be sure to only count regions moving as a unit. Finally, we used HMMER (*63*) to search each element for known defense genes using Supplementary Tables 1 and 5 in Doron et al. (*15*) and Supplementary Tables 1 and 8 in Gao et al. (*17*). FASTA files for all regions are provided in Data S8.

Gene annotations of the PDEs were performed manually using the consensus obtained from HHPred (*82*), InterProScan5 (*62*), Phyre2 (*56*), and BLASTp (*83*) databases tools (Data S3). The search was performed using default options except that HHPred search was performed against COG-KOG 1.0 and Pfam-A_v32.0 databases and that BLASTp was performed using the protein-protein BLAST option. Up to ten significant pfam and COG (p < 0.05) from HHPred search where used to compare each gene with pfam-COG accession numbers of phage defense systems from Supplementary Tables 1 and 5 in Doron et al. (*15*) and Supplementary Tables 1 and 8 in Gao et al. (*17*).

Finally, to determine the relationship among the PDEs (Fig. 1C), we clustered the proteins on the PDEs by 50% identity using CD-HIT (*84*) and visualized the network using Gephi (*81*). This low identity threshold only allowed for a limited number of connections, indicating the diversity and unrelatedness of the elements.

Figure 1 depicts a closely related set of 20 strains, which range in divergence from 0-95 SNPs, while on average 757 SNPs differentiate any one of the 20 from the 2 slightly divergent strains and 16,953 SNPs differentiate any one of the 22 from the outgroup. This collection of hosts was essential for the comparative genomics determining PDE turnover. The approach we present here: using comparative genomics and genetics of near clonal isolates combined with phage host-range data is a fruitful method to discover novel phage defense mechanisms in an unbiased way, in addition to monitoring all modes of evolution in response to phage predation (SNP resolution as well as HGT). However, we emphasize that near clonal genomes of high quality are essential since in our experience PDEs assemble poorly when using short read data, in part due to AT richness and high density of variable repeat regions.

Transfer of phage defense elements

To further characterize the distribution of the 5 specific PDEs identified among the 19 "orange" and "purple" strains in a larger collection of *Vibrio* genomes (*37*), we used BLASTn (*64*) and custom python scripts to identify the distribution of the mobile elements (Fig. S9, Fig. S10). Because when comparing long mobile elements BLASTn will often return multiple overlapping ranges of identity, our BLAST parsing script merges overlapping sequence ranges to avoid over-counting regions within a genome. A PDE was considered present in a genome if at least 80% of the element was present at over 95% identity.

Identification of insertion sites

To determine the genomic context for each PDE and find the same region in genomes without the PDE, genes identified using prodigal 2.6 (*71*) were clustered at 80% identity using mmseqs2 (*60*). For each representative region, we searched for the closest single copy core gene upstream and downstream of the representative PDE. Then for each strain where these core genes were present on the same contig we plotted the intervening genes (Fig. S6).

Proportion of known defense genes in flexible genome across diverse *Vibrio*

To determine the proportion of known defense genes in other *Vibrio* species, we used the species and population designations from Arevalo et al., (*37*). We based our identification of flexible genes on the method described in Arevalo et al. ORFs were identified with Prodigal 2.6 (*71*) and orthologous genes were clustered using MMseqs2 (*60*). Flexible genes for a given population were defined as orthologs which were present in at least one member but not present in all members of the population. Flexible genes from each genome were then used as a database which we searched for known defense genes using Supplementary Tables 1 and 5 in Doron et al. (*15*) and Supplementary Tables 1 and 8 in Gao et al. (*17*) using HMMER (*63*). Total length of all flexible genes summed for each species and the proportion of genes with a hit to a known defense gene is shown in Figure S7.

Identification of putative PDEs from comparison of closely related genomes of different bacteria

We extended the search for novel putative PDEs by searching identify nearly clonal genomes of *Salmonella*, *Listeria,* and *Clostridia*. For *Salmonella*, and *Listeria* we selected genomes within

the same ribotype using the ribosomal MLST database (*62*), filtered to only include NCBI assemblies, and downloaded genomes from ribotypes with more than 20 members. We used ribotype 8354 for *Salmonella* and MLST strain type 5 for *Listeria* to assay putative PDE distribution in an exemplary manner. For *Clostridia*, we focused on a set of clones with identical ribosomal proteins we hand selected due to limitations in available genome quality. In all cases, the final set of genomes was run though the custom kmer-based comparative genomic pipeline to identify flexible regions (Data S2). All pairs of genomes were compared (see "Identification and annotation of putative PDEs in the flexible genome" for details). As with the *Vibrio* analysis, we removed any element that appeared in all genomes and used HMMER (*63*) to search each element for known defense genes using Supplementary Tables 1 and 5 in Doron et al. (*15*) and Supplementary Tables 1 and 8 in Gao et al. (*17*). Any unique region with one or more hits was considered to be a putative PDE and depicted in Figure S20, Figure S21, and Figure S22 for *Listeria*, *Salmonella* and *Clostridium*, respectively. Additional annotations were performed with EggNog-Mapper (*57, 58*). FASTA files for all regions are provided in Data S8.

We faced the same issue of genome completion explained above when looking at other vibrio populations as well as the *Listeria*, *Salmonella*, and *Clostridia* analyses, but tried to minimize these issues by looking at closely related (identical ribosomal protein sequences) strains and using the highest quality genomes available.

Proportion of the flexible genome attributed to PDEs

To determine the proportion of the flexible genome attributed to PDEs in *Vibrio*, *Salmonella*, *Listeria,* and *Clostridia*, we used the PDE analysis presented for each genus as an exemplar and determined how many genes were attributed to the identified PDEs compared to the number of genes on other flexible regions identified in the k-mer pipeline. The breakdown is included as a separate tab for each group in Data S3-S6.

Methylome characterization and restriction-modification activity

To discover if the restriction modification systems identified on the PDEs in the "orange" and "purple" strains are active, we determined the methylation sites in both phage-host pairs as outlined in Murray et al. (*85*). Briefly, we submitted the host genomes to REBASE, a well-curated database of restriction modification systems which allows motif prediction based on comparisons to known enzyme-motif pairs (*86*). Then, we combined the motif data with the methylome data generated using the Base Modification Detection and Motif Analysis pipelines available on the single molecule real-time (SMRT) sequencing portal from Pacific Biosciences. Summary data is presented in Table S4.

Methylome sequencing analysis shows that the distinct sequence motifs predicted for each RM system using REBASE are methylated only within "orange" and "purple" phage-host sets, i.e., only in the bacterial genomes in which the RMs occur, and in the phages that can kill those hosts (Table S4).

To further characterize the escape phenotype of the phages resulting in their ability to kill, we propagated a "purple" phage grown on the triple knockout "orange" host and measured the efficiency of plating on the WT "purple" host (Fig. S14). We saw a reduction of efficiency of plating by seven-fold, yet propagation of plaques was still observed. This is the expected phenotype for a methylated escape phage mutant. Only phages that are inadvertently methylated will be able to kill the WT host, while this methylation is not needed for the triple knock out as it is lacking the corresponding RM system on PDE1.

Bacterial strain selection for transposon mutagenesis and gene deletions
Because the adsorption assays indicated that both "orange" and "purple" phages adsorbed to both host groups, we chose one strain, 10N26155C8 (Orange WT, hereafter C8-WT), for mapping of receptors for both host groups. For receptor mapping, we took advantage of the lysis from without phenotype where phages can affect lysis if hosts possess a specific receptor even if no viable phage are produced. Accordingly, at high phage titer, cells of both host groups are lysed by both phages (Fig. S3), allowing for testing of receptors using a "purple" phage on an "orange" host. The same C8-WT strain was used for characterization of resistance determinants of the "orange" host group by gene deletion (see below).

Growth conditions of strains used for transposon mutagenesis and gene deletions
C8-WT was routinely grown at 25ºC in 2216MB or TSB2. The *Escherichia coli* strains were grown in BD Difco Miller Luria-Bertani broth (LB) at 37ºC and supplemented for auxotroph strain *E. coli* Π3813 with thymidine (0.3 mM), and for strains *E. coli* β3914 and MFDpir with diaminopimelic acid (dapA) (0.3 mM). Antibiotics were used at the following concentrations: erythromycin (Erm) 200 μg mL^{-1}, kanamycin (Km) 50 μg mL^{-1} and chloramphenicol (Cm) at 5 or 25 μg mL^{-1} for *Vibrio* and *E. coli*, respectively.

Receptor identification
Transposon mutagenesis as well as resequencing of spontaneously resistant strains all suggest that the "orange" phage receptor is the Type II secretion system pseudopilus (GspH), while the "purple" phages have two receptors, one being the lipopolysaccharide (LPS), an extracellular receptor, and the second being a sodium transporter (NqrC), a membrane receptor (Fig. S5A, Tables S1 and S2). These results are further verified by clean deletions and/or complementation (Fig. S5BC).

Receptor identification using transposon mutagenesis
To map phage receptors, transposon mutagenesis was carried out using suicide delivery of a mariner transposon. C8-WT served as the recipient and the dapA deficient strain *E. coli* MFDpir with the suicide conjugative plasmid pSC189-Cm (*87*) served as the donor (Data S7). The delivery plasmid (pSC189) can be mobilized via RP4-mediated transfer and it carries the hyperactive C9 mariner transposase (*88*). Conjugation was carried out by mating assays as described previously (*89*) with some modifications. First, donor:recipient ratio was adjusted to 1:3. Overnight cultures were diluted 1:100 in fresh media and grown up to an OD$_{600}$ of ~0.4. One mL was separately pelleted at 5,500 x *g* for two minutes and washed in pre-warmed mating media broth MMB-1 (TSB supplemented with 1% NaCl plus dapA) to remove antibiotics and/or residual media. This wash step was repeated twice. Washed pellets were subsequently mixed in the same tube with 500 μL of MMB-1, pelleted and resuspended in a mating spot (20 μL) on mating media agar plates and incubated at 25ºC for 18 hours. Mating spots were collected using a Nunc 10 μL sterile plastic inoculation loop and resuspended in 500 μL of ASW. Then, 100 μL of this suspension were spread onto TSB2 plates supplemented with Cm and incubated at 25ºC for 48 hours. Finally, the mutant library (totaling 26,662 mutants) was archived in 500 μL aliquots with ASW supplemented with Cm and 25% glycerol (v/v), quickly frozen in a dry ice bath for 10 minutes and then stored at -80ºC until testing.

365   Resistant mutants were selected by challenging the mutant library with high titers of phages. Four aliquots of the mutant library were defrosted, centrifuged by pelleting at 5,000 x *g* for five minutes, and then washed twice with 2216MB to remove any residual glycerol. The washed pellets were then resuspended in their original tube with one mL of fresh 2216MB supplemented with Cm. C8-WT served as positive control and was treated equivalently, except

370   for the addition of Cm. The washed mutant library and C8-WT control were both diluted 1:10 in 2216MB and then incubated at room temperature with vigorous shaking (250 rpm) for 1 hour until the cultures reached early exponential phase. To select for phage-resistant mutants, lysates was serially diluted 10-fold and mixed with the mutant library and C8-WT cultures. Aliquots of host-phage culture were mixed into 750 μL of 2216MB top agar (with and without Cm as

375   needed) and spread on large 2216MB bottom agar plates (with and without Cm as needed) following the soft agar protocol as described previously (*54*). After incubating at room temperature for 48 hours, ~100 phage-resistant colonies were selected at random and serially re-streaked three times on 2216MB agar plates (with and without Cm as needed). Glycerol stocks of each mutant were archived, and all mutants were then re-tested for phage susceptibility. The re-

380   test was always done with two "orange" and two "purple" phages, one of each always being the original phage used to isolate resistant colonies. In all cases, resistance to one "orange" phage yielded resistance to all "orange" phages and resistance to one "purple" phage yielded resistance to all "purple" phages. Cross resistance to opposite or both phage groups was never seen, further supporting the finding that each group of phages uses a different receptor.

385   Arbitrary PCR (*90*) was used to map the transposon insertions in resistant strains. Genomic DNA from each phage-resistant mutant was extracted with Lyse-n-Go direct PCR reagent (Thermo Scientific), and one μL of the lysate served as template in arbitrary PCR. This method involved two rounds of PCR amplification (*90*): in the first round, genomic DNA was amplified with a fully degenerate primer SS9arb2 (Data S8) containing a 5′ tail of known-

390   sequence to be used for specific amplification in the second round of PCR (*91*), paired with primer Mar4 (Data S8) that binds the end of the transposon TnSC189 (*92*). Optimized conditions for the first round PCR consisted of the following reagent concentrations and amplification parameters: primers SS9arb2 and Mar4 were at 0.5 mM and 0.2 mM, respectively; GoTaq G2 HotStart (Promega) was used with $MgCl_2$ at two mM; initial heating for two minutes at 95°C,

395   followed by six cycles of 30 seconds at 95°C, 30 seconds at 30°C, and 1 minute and 30 seconds at 72°C; 30 cycles of 30 seconds at 95°C, 30 seconds at 55°C and one minute and 30 seconds at 72°C, with a final extension for five minutes at 72ºC. In the second round of PCR amplification, 2.5 μL of the first-round PCR product was used as template, combined with a nested primer within the amplified fragment of TnSC189 (Mar4_int2) and primer (Arb3) with sequence

400   identity to the 5' tail of the SS9arb2 (Data S8). For the second round, PCR reagents were used as described above but using both primer concentrations were 0.2 mM, and the PCR was run under the following conditions: two minutes at 95°C, 30 cycles of 30 seconds at 95°C, 30 seconds at 58°C and one minute and 30 seconds at 72°C, with a final extension time of five minutes at 72ºC. PCR products were verified by electrophoresis, purified by spin-column using QIAquick

405   PCR Purification (Qiagen) and then Sanger-sequenced. Finally, amplicons were trimmed and mapped to the C8-WT genome to identify transposon insertion locations using a custom python script and BLASTn (*64*). Hits are present in Table S1.

Receptor identification using re-sequencing of spontaneously resistant isolates

As an independent method to transposon mutagenesis, we re-sequenced spontaneously resistant mutants from co-cultures of "orange" host C8-WT and high titer phages to identify phage receptors. C8-WT was streaked out from glycerol stock onto 2216MB agar plates, inoculated into five mL of 2216MB liquid media, gown shaking overnight at room temperature, and plated as a lawn in a soft agar overlay. Five µL drop spots of a phage dilution series were plated on top of the agar, and after 24 hours, resistant colonies that grew in the presence of high phage concentrations were re-streaked three times and archived. For each phage, 10 colonies were archived. Resistant strains were re-streaked and re-tested to verify resistance. The re-test was always done with two "orange" and two "purple" phages, one of each always being the original phage used to isolate resistant colonies just as in the transposon experiments. The results were also consistent with the transposon mutagenesis experiments: In all cases, resistance to one "orange" phage yielded resistance to all "orange" phages and resistance to one "purple" phage yielded resistance to all "purple" phages. Cross resistance to opposite or both phage groups was never seen. Six to seven strains verified in this way were sequenced on an Illumina HiSeq as described in the "Hybrid genome assemblies" section. Reads were trimmed and mapped to the hybrid assembly reference genome using CLC work bench 9. Single nucleotide polymorphisms (SNPs) and indels were identified using a custom pipeline made for CLC work bench 9 and are presented in Table S2. These SNPs were cross-referenced to receptor identification using transposon mutagenesis (see above).

Receptor verification using gene deletion or complementation
To formally demonstrate the role of a gene predicted to encode a phage receptor, clean deletions were performed using ccdB based suicide vector pSW7848T (*89, 93*). Fragments (~0.5 kbp) flanking the locus to delete were amplified and assembled by PCR, then cloned into pSW7848T using the Gibson Assembly Master Mix (New England Biolabs) according to the manufacturer's protocol. The plasmid was transferred from the *E. coli* donor strain (u3914) to the recipient vibrio by conjugation as previously described. Selection of the plasmid-borne drug marker on TSB-2 supplemented with Cm 5ug/ml and glucose 1% resulted from integration of pSW7848T in the genome. The second recombination leading to pSW7848T elimination was selected on TSB-2 supplemented with arabinose 0.2%. Mutants were screened by PCR using external primers flanking the different targets.

 Because GspH seemed to be essential, and we could not knock it out, we took the spontaneously resistant mutant with a SNP in *gspM* and complemented in the WT of the same gene to restore killing. The WT gene was amplified by PCR (see primer details in Data S8) and cloned in Apa1/Xho1 sites of the pMRB plasmid under the control of a constitutive Plac promoter (*94*).

Distribution of putative receptors across diverse vibrios
To test how diverse receptor genes identified by mutant analysis in *V. lentus* are across other vibrios, we used the species and population designations from Arevalo et al., (*37*). In cases where a species was composed of multiple populations, we chose to only analyze the population with the most members. We based our identification of core genes on the method described in Arevalo et al. (*37*). ORFs were identified with Prodigal 2.6 (*71*) and orthologous genes were clustered using MMseqs2 (*60*). Core genes for a given population were defined as orthologs which were present in a single copy in all members of that population. We aligned all genes within orthologous clusters using MUSCLE (*95*) and calculated the amino acid diversity between

each pair of genes as the number of positions with non-identical amino acids divided by the alignment length. Average pairwise amino acid diversity was obtained by taking the mean diversity across all pairs of genes within an orthologous cluster. The amino acid sequences of each receptor gene were aligned using mafft-linsi (*74*) and average pairwise amino acid diversity was calculated as described above.

Phylogenetic relationships among receptors identified using genetic approaches were determined across a collection of diverse *Vibrio* isolates (*73*). Receptor orthologues were identified using BLASTn (*83*) with evalue of 1E-8, aligned using MAFFT (*74*), and final trees were constructed using RAxML (*75*) with model GTRCAT.

To estimate whether diverse isolates are decorated with similar lipopolysaccharides (LPS), we extracted the genome region bounded by *gmp* and *gmhD* (*96*), which encodes genes for LPS synthesis. We searched the same *Vibrio* genomes from Arevalo et al. (*37*), that were used for the determination of known defense genes in Figure S7. When the same set of genes was present, we inferred that the LPS structure was highly similar and may thus bind to similar phage receptors. We only analyzed genomes that had LPS regions that were present on a single contig to avoid ambiguity in presence/absence patterns. Genes were clustered using MMseqs2 (*60*) at 90% identity and required to be present in at least 2 genomes. Presence/absence profiles are shown in Figure S18C using hierarchical clustering to order the genomes (rows) and LPS genes (columns). Genes with identical presence absence patterns in Figure S18C have only one tree branch going into the left-most gene (top tree).

RNA-Seq analysis of "orange" PDEs
We diluted 200uL an overnight culture of the WT "orange" host in 200mL of fresh media and grew it to mid exponential phase in an Erlenmeyer flask. Then, we added a representative "purple" phage at an MOI=10 or MOI=5. Every 15 minutes, for a total of 11 timepoints, we archived 500uL of the culture in Qiagen RNAProtect, and then placed on dry ice. Once the time course was collected, we extracted the total RNA for sequencing using a Qiagen RNEasy extraction kit, following the manufacture's instructions. RNA was sequenced at the BioMicro Center at MIT. By mapping the transcripts to the host genome, we were able to identify regions that appeared to be transcribed together. Reads were filtered and trimmed using trimgalore version 0.5.0 with the following options: trim_galore -e 0.05 --length 70 --max_n 1 --error_rate 0.05 --stringency 1 --clip_R1 1 --three_prime_clip_R1 1. Reads originating from the virus and from ribosomal RNAs were removed using bbslit version 38.18. Reads were mapped to the region using bbwrap version 38.18 with the strandedcov=t option to later manually identify putative transcriptional units. Gene expression was estimated using Kallisto version 0.46.2 with the following options: kallisto quant --single -l 200 -s 20. Both replicates (MOI=10 and MOI=5) were consistent in expression for the 3 PDEs analyzed (Fig. S11).

Phage defense element knockouts using two-step allelic exchange
To test whether putative PDEs were responsible for phage resistance in C8-WT, we knocked out large portions of each PDE containing genes annotated as being related to phage defense (Fig. 2A), in all possible combinations. We found it was possible to knock out nearly all of PDE1 (93.5%), but had to leave in part of the element's *yfbr* gene as it replaces the host's *yfbr* gene upon insertion. For PDE2, we found that knocking out the entire element was not possible in a single step, likely because of the toxicity effects of deleting the entire putative TA/Abi systems at once. We therefore proceeded to make a partial deletion (58.8%) from the helicase/toxin gene to

the 5' end of the element, leaving the putative antitoxin intact, along with other genes predicted to play roles in insertion/mobilization of the element (integrases, transposases, and recombinases) as indicated by structure and function annotations using HHpred (*97*), InterProScan5 (*62*), and Phyre2 (*56*) (Fig. 2A, Data S3). Noting PDE3 also has a similar putative TA system, we followed the same approach as for PDE2 and made a partial deletion (68.8%) from the helicase/toxin gene to the 5' end of the element. The details of the approach are summarized in Figure 2A.

Site-directed mutagenesis was used for all deletions and cloning was carried out using the New England Biolabs Gibson Assembly Master Mix according to the manufacturer's protocol. Fragments upstream and downstream of the portion of the element to be deleted were separately PCR-amplified using primers specified in Data S7. A third PCR reaction was carried out to amplify the backbone of the plasmid pSW7848T (*93*) (Data S7) with primer pairs pSW_F&R (Data S7). These amplicons were cut with Dpn1 (two hours at 37ºC) to inactivate the plasmid template before setting up the Gibson assembly reaction. In all cases, PCR products were verified by gel electrophoresis, purified by spin-column as described above, and DNA concentration was determined using Nanodrop 2000 (Thermo Scientific). Subsequently, 0.03 pmol of each vector was assembled with 0.07 pmol of its specific downstream and upstream DNA fragments at 50ºC for 60 minutes. After completion of this reaction, DNA was desalted by dialysis on a 0.0025 µM filter (Millipore) before electroporation into *E. coli* Π3813, which was used as a plasmid host for cloning (*89*). Finally, the plasmid DNA was purified, verified by Sanger sequencing, and electroporated into *E. coli* β3914 to be uses as a plasmid host for conjugation (*89*) (Data S7).

Conjugation was carried out in a mating spot as described above for the transposon mutagenesis but with some modifications: donor to recipient ratio was changed to 3:1, the mating media was altered to MMB-2 (TSB supplemented with 2% NaCl plus dapA), and the mating spot was incubated at 30ºC. Counter-selection of ΔdapA donor was performed by plating on TSB2 agar plates without dapA but supplemented with Cm and glucose 1% (w/v). Antibiotic-resistant colonies are due to the integration of the entire plasmid (Cm^R) in the chromosome by a single crossover. Colonies were picked, re-grown in liquid media (TSB2) supplemented with Cm and glucose 1% (w/v) to late logarithmic phase and spread on BD Bacto TSB without Dextrose plates supplemented with 2% NaCl (w/v) and 0.2% arabinose. To verify deletions in the single PDE mutants ΔPDE1, ΔPDE2 and ΔPDE3 (Data S7) PCR products generated using primers flanking externally the different regions targeted (ΔPDE1/F&R; ΔPDE2/F&R and ΔPDE3/F&R) (Data S7) were sequenced by Sanger. This procedure was also used to construct double (ΔΔPDE12; ΔΔPDE13; ΔΔPDE23) and triple mutants (ΔΔΔPDE123) but using a single or double mutant as final recipient during the conjugation step (Data S7).

Single phage defense gene characterization using transposon mutagenesis or complementation
To identify the specific genes involved in phage defense and further characterize the PDEs discovered in this work, we took advantage of our double mutants to scrutinize the third remaining PDE. For example, using the ΔΔPDE23 knockout, we used integration of a suicide vector by single recombination to inactivate genes hypothesized to underlie phage defense (using the same protocol described under "receptor verification"). Using this approach, we identified that inactivating the restriction gene of the restriction modification system on PDE1 was necessary to yield loss of defense. The analogous approach was also successful in determining the genes involved in phage defense on PDE2. We found four genes within a module on the 5'

end of the element was each involved in defense: unannotated genes 10N26155C8 ORF_0_597, putative GTPase 10N26155C8 ORF_0_598, putative GTPase 10N26155C8 ORF_0_599, unannotated gene 10N26155C8 ORF_0_600, and unannotated gene 10N26155C8 ORF_0_601 with DUF4231.

However, this method was unsuccessful for determining the genes involved in PDE3. While we used the same approach to inactivate 11 different genes, which we hypothesized to be involved in defense based on superficial annotations, no genes responsible for defense to "purple" phages were identified. Instead, we took a transcriptomics approach (described under "RNA-Seq of "orange" PDEs"). This allowed identification of 14 regions that appeared to be transcribed together. All regions were cloned by Gibson assembly in a pMRB plasmid previously depleted from the Plac promoter by a PCR inside out. These regions carry 2-3 genes under the control of native promoters. Each construct was transferred into the C8 triple mutant ΔΔΔPDE123 that is fully sensitive to the "purple" phage. Although six regions were not able to be cloned into *E. coli*, possibly due to toxicity, and thus could not be further analyzed (R1, R2, R4, R5, R11, R13), this approach led us to identify Region 3 as sufficient for defense. This region contains only two predicted full length ORFs: ORF 806, which is a short unannotated gene, and ORF 807 which is annotated as having a conserved domain of unknown function (DUF2971). However, through inactivation, we did not see any phenotype for ORF 806, leading us to conclude ORF 807 was most likely the gene of interest.

Phage susceptibility assay for whole PDE deletion mutants

To test the susceptibility of the "orange" PDE deletion mutants to "purple" phages, we challenged each mutant with representative "purple" phage 1.281.O in agar overlays (Fig. 2B) and in liquid culture (Fig. S13). The plate-based findings were corroborated in liquid, although the subtle phenotypic alterations for single PDE knockouts required careful plating to observe. These experiments were also repeated to test the mutants against all "orange" and "purple" phages as well (Fig. S12).

For the mutant testing in agar overlays, we used the same protocol outlined in the "Host range assays at varying phage concentrations" section above, with one additional step: after the plaques were imaged, we re-streaked the phages from the highest concentration drop spot onto fresh bacterial (host or mutant) lawns to test for phage propagation (Fig. 2C). We imaged the agar overlay plates using a flatbed scanner (Epson Perfection V800 Photo Scanner - Product No. B11B22320) and captured using VueScan Software by Hamrick as before. However, for the entire panel of data in figures 2B and 2C alone, we also optimized visualization by manipulating the images in Adobe Photoshop. Specifically, applying "Auto contrast" on all images. No other images throughout the text were manipulated.

For the liquid assay, we streaked out each mutant onto 2216MB agar plates, allowed 48 hours for large colonies to form, and inoculated each mutant into 3 mL of 2216MB in triplicate. After growing the cultures shaking at 25°C for four hours, we normalized one mL of the culture to an OD_{600} of 0.3, diluted it 100x into a final volume of five mL, and aliquoted 200 μL into 12 wells each of a 96-well clear bottom Micro-titer plate (Falcon). A Tecan Microplate Reader with Spark software was used to maintain the cultures shaking at 25°C, monitoring OD_{600} every 15 minutes. Once OD_{600} reached 0.3, "purple" phage 1.281.O was added at different concentrations to reach the desired multiplicities of infection (Fig. S13), after which the cultures were returned to the plate reader for the remainder of a 24-hour run. This experiment was run in triplicate for

each biological replicate. It was run with two mutants, C8-WT, and "purple" host 10N28654F7 each time until all mutants had been tested.

595

Quantification of PDEs within clonal hosts across the Nahant Time Series

Samples for total community DNA were described previously (*36*). Briefly, samples were collected on each of the 94 days in the time series. Upon return to the lab, water samples were filtered using a peristaltic pump through a 0.2um Sterivex filter into a secondary 4L collection

600 bottle. To extract total community DNA, filters were carefully cut and placed in two mL tubes with 0.25 g of sterile 0.1 mm zirconium beads and 750 µL of Qiagen Cell Lysis Solution, and then run on a Beadbeater at 5000 rpm for 60 seconds. Samples were then incubation at 80 °C for five minutes and cooled to room temperature, and RNA was digested by adding four µL RNAse A (four mg/mL), mixed by inverting the tubes, and incubated at 37 °C for 30 minutes. DNA was

605 purified by adding 250 µL of Qiagen Protein Precipitation Solution, mixed by vortexing for 20 seconds, incubated on ice for five minutes and centrifuged at 13,000 rpm for five minutes. Supernatant was transferred to a new tube and centrifuged again to ensure removal of all precipitates. Subsequently, 750 µL of isopropanol were added to 750 µL of supernatant to precipitate DNA and incubated at −20 °C overnight. DNA was recovered by centrifugation at

610 13,000 rpm for five minutes and washed with 700 µL of 70% ethanol. Finally, the DNA was dried and resuspended in 100 µL of Qiagen brand DNA hydration solution.

To quantify PDEs carried within a specific host genomic background, we carried out digital droplet PCR (ddPCR) on the time series total DNA samples. Custom ddPCR assays were designed using Primer3Plus with the "Probe" server settings activated to target three inserted

615 phage defense elements, two unique to the "orange" group (PDE2 and PDE3) and one unique to the "purple" group (PDE4). The amplified product always spanned the insertion site of the element, while the probe targeted either the insertion site or the element of interest depending on design constraints of the assay. Complementary probes (FAM/HEX) were used to allow for assay multiplexing and final assays were ordered through Bio-Rad. For the PDE2 and PDE4

620 multiplexed assay, 20 µL of a 22 µL reaction containing 11 µL ddPCR Supermix for Probes (No dUTP) #1863024 master mix, 1.1 µL PDE4 Assay, 1.1 µL PDE2 Assay, 1.1 µL BsmI restriction enzyme (NEB R0134L) diluted in CutSmart buffer, and 7.7 µL of 10-fold diluted sample were prepared for analysis and run on the standard ddPCR thermocycler program using a Tm of 60°C. For the PDE3 assay, 20 µL of a 24 µL reaction was run using standard recommendations from

625 Bio-Rad. All primers and probes are listed in Data S7.

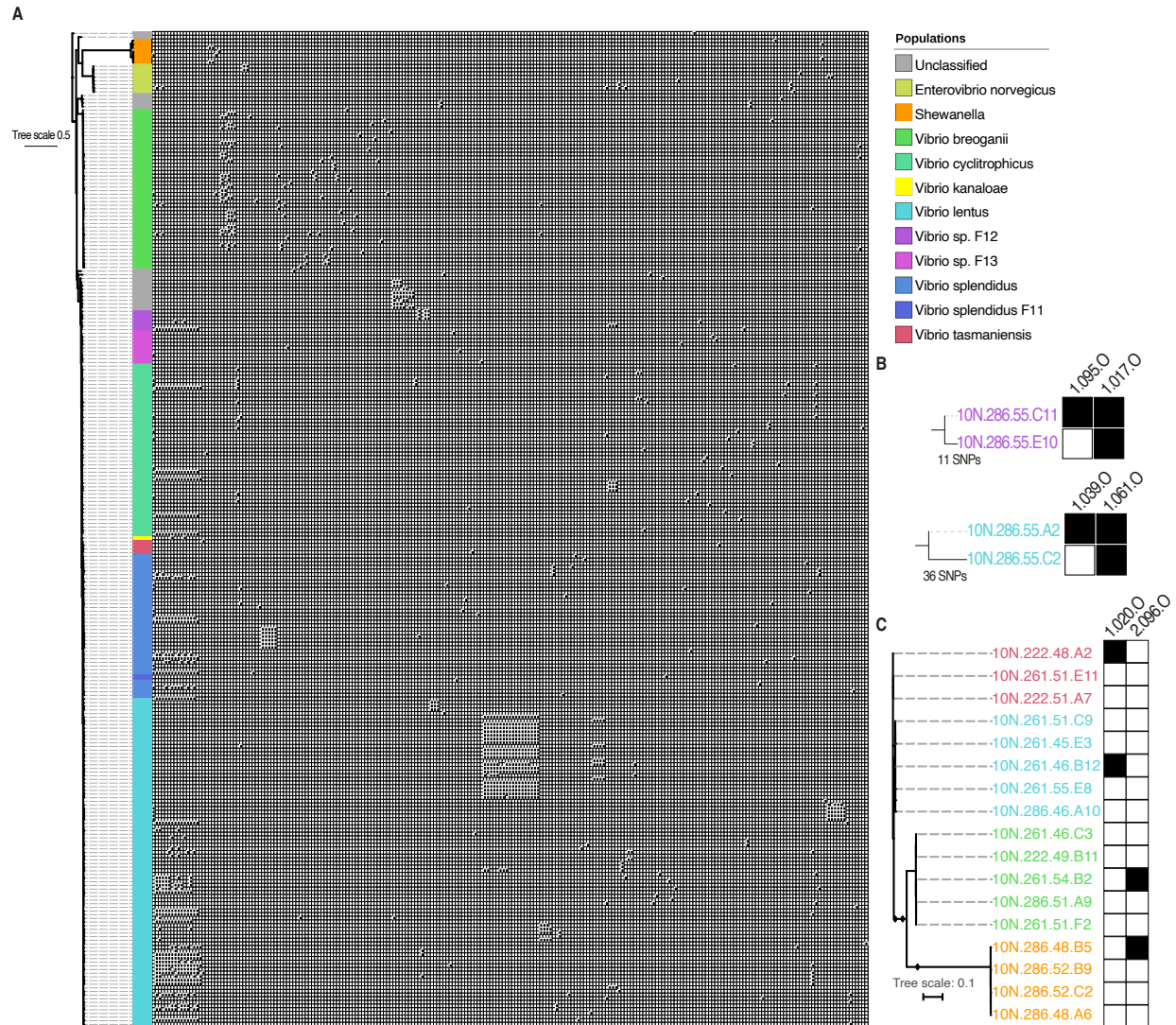Fig. S1.

Phage host-range established using an exhaustive cross-test matrix. (A) Full matrix with rows
depicting bacterial hosts organized by the phylogeny of their ribosomal protein and *hsp60* gene
sequences (proxy for core genome), and columns depicting phages ordered by protein similarity
identity [modified from Figure 2 in (*34*)]. (B) Closest bacterial relatives differing in phage
sensitivity profiles can be distinguished by only few SNPs across their entire core genomes.
Trees represent full genome alignments, phage identification codes written above columns, black
boxes indicate positive infection determined by plaque assay. (C) Broad host-range phages,
defined as host ranges spanning different species, remain strain-specific within different species.
Phylogenetic tree constructed using same alignment of core genes as in A, and infection
representation analogous to that in B. Source data available in Kauffman et al. 2021 (*35*).
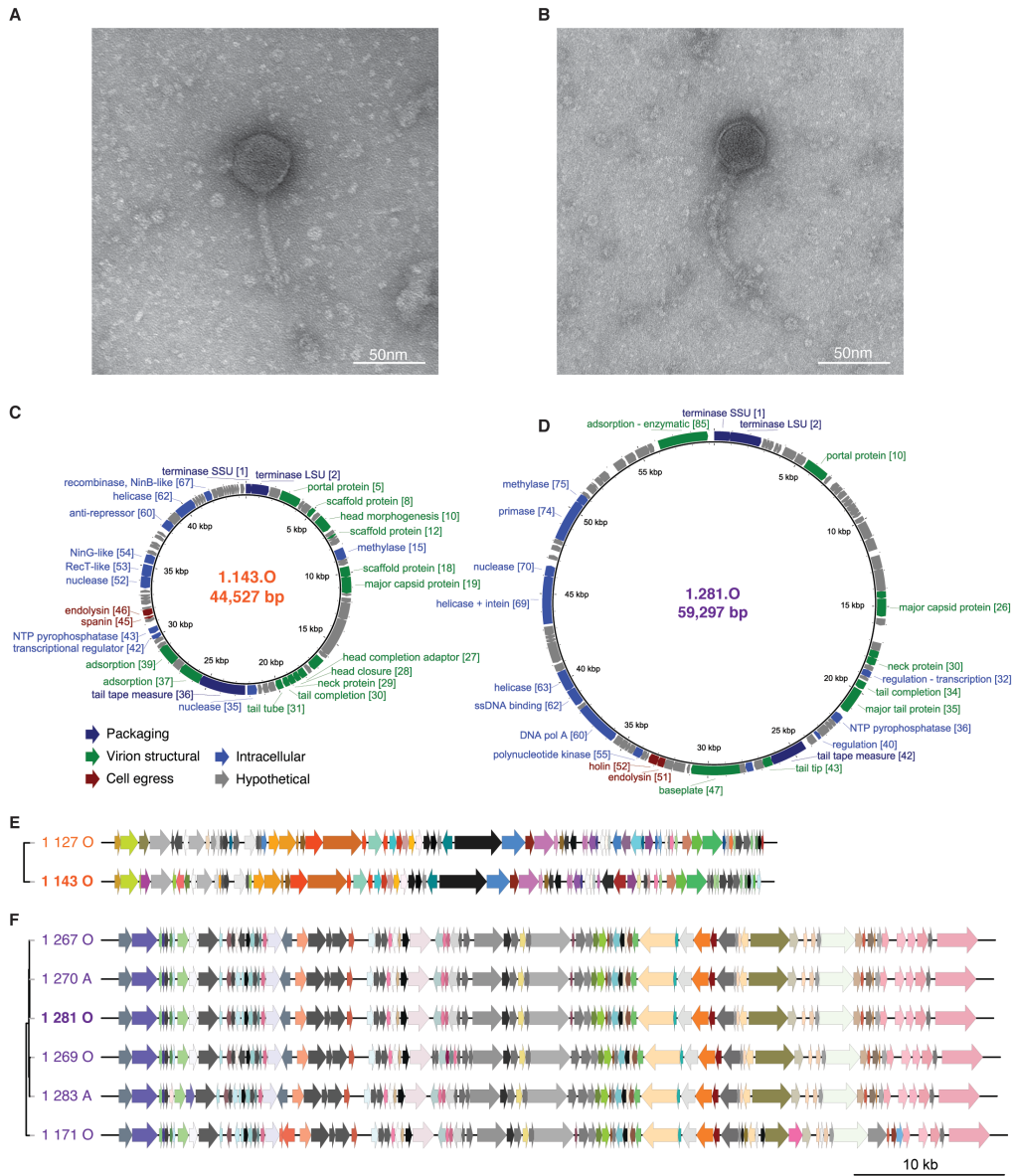
Fig. S2.
"Orange" and "purple" phages represent divergent groups of siphoviruses. (A, B) Electron microscopy of phages representative of "orange" (MG592514.1) and "purple" (MG592649.1) groups suggests that both are siphoviridae, with long non-contractile tails. (C, D) Genome characterization of phages representative of "orange" and "purple" groups, respectively, shows that they differ in size by nearly 15 kbp; numbers adjacent to annotations reflect GenBank locus tag. (E, F) Clustering and alignment of phage genomes show that they represent two distinct genus-level groupings. While within each group gene synteny and content are conserved, no gene clusters are shared between groups. Colors are chosen at random to indicate distinct protein sequence clusters (clustered using MMseqs2 v. 2.23394 under default settings). Accessions of phages shown: 1.127.O, MG592501.1; 1.143.O, MG592514.1; 1.267.O, MG592632.1; 1.270.A, MG592636.1; 1.281.O, MG592649.1; 1.269.O, MG592635.1; 1.283.A, MG592651.1; 1.171.O, MG592538.1.

A

1:10 dilution series of representative "orange" phage 1.143.O

"Orange" host representative 10N.261.55.C8

"Purple" host representative 10N.286.54.F7

B

1:10 dilution series of representative "purple" phage 1.281.O

"Orange" host representative 10N.261.55.C8

"Purple" host representative 10N.286.54.F7
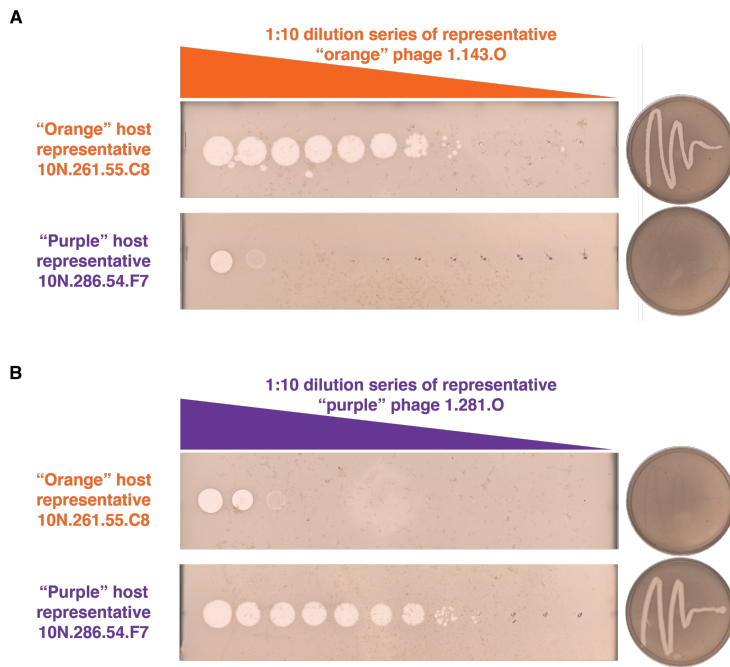
Fig. S3.

Efficiency of plating assay demonstrating effect of differing phage concentrations on host killing. At high concentrations, phage can lyse non-hosts but without production of viable progeny ("lysis from without") indicating that phage can attach and enter the cell, but that replication is prevented internally. (A) "Orange" and "Purple" host lawns challenged with serial dilutions of "orange" phage followed by re-streak propagation test on the right. (B) "Orange" and "Purple" host lawns challenged with serial dilutions of "purple" phage followed by re-streak propagation test on the right.
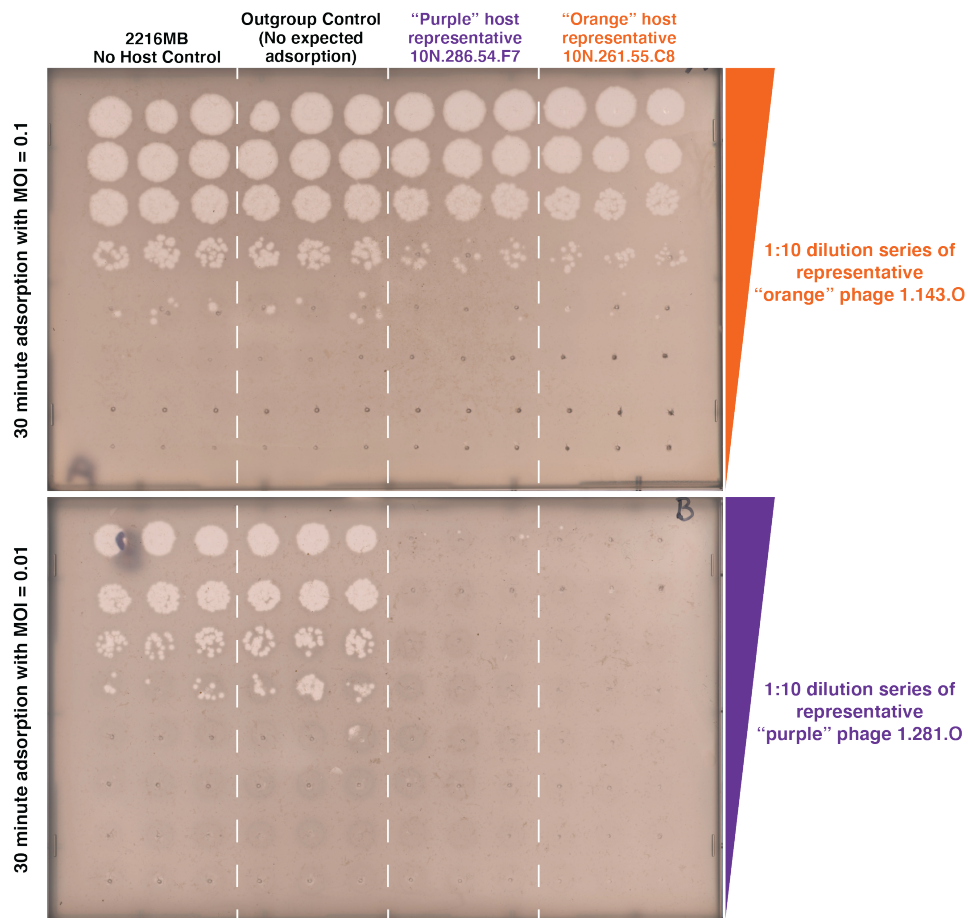
Fig. S4.

Phage adsorption assay showing that phage can adsorb to both "orange" and "purple" strains irrespective of whether those bacterial strains can serve as hosts for viable phage production. After allowing a fixed concentration of phages to adsorb to different bacterial strains, free phages that remained unattached were plated with sensitive hosts to quantify adsorption as the difference to no-host controls (see methods). Both "orange" and "purple" phages were found to adsorb to "orange" and "purple" hosts, but not to an outgroup control. In the top panel, "orange" phage 1.143.O shows the same adsorption phenotype to both "orange" host 10N26155C8 and "purple" host 10N28654F7: the number of free phages decreased by ten-fold. In the bottom panel, "purple" phage 1.281.O shows the same adsorption phenotype to both "orange" host 10N26155C8 and "purple" host 10N28654F7, attaching with full efficiency. In both cases, no attachment is observed for a *Vibrio* outgroup host (10N26149C11) as indicated by the same level of phages as in no host controls. Adsorption experiment done with biological and technical triplicates of the represented "orange" and "purple" phages and hosts in this figure.

18
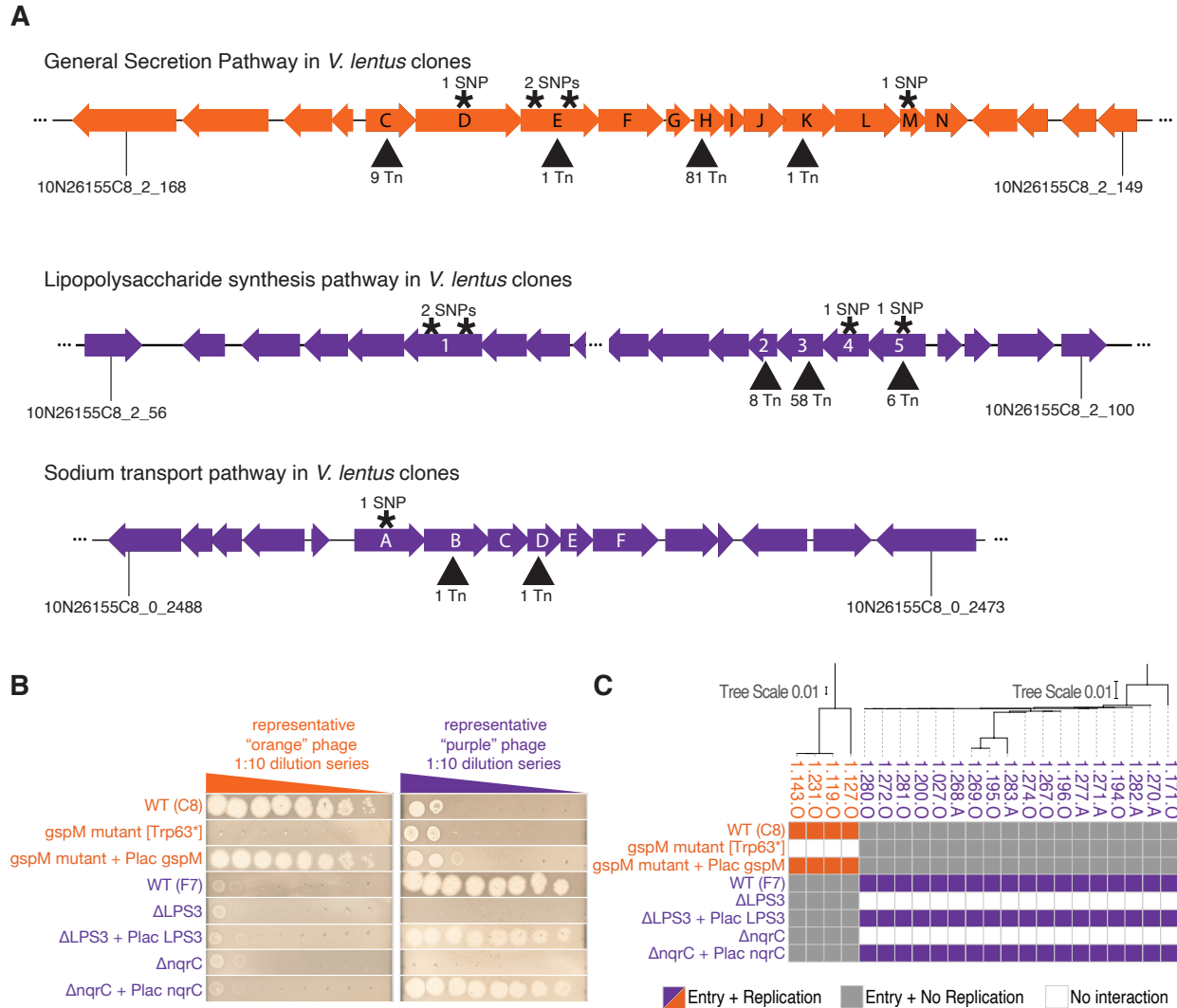
Fig. S5. Receptor identification using molecular genetics and experimental evolution. (A) Gene diagrams of loci associated with each receptor in representative "orange" strain 10N26155C8. Transposon insertions depicted with black triangles and SNPs in spontaneously resistant clones indicated with *. (B) Representative lawns of each strain with either SNP, deletion, or complementation challenged with a dilution series of a representative "orange" and "purple" phage. (C) Graphical summary of all-by-all testing of receptor mutants with all phages tested in triplicate (example in B, summarized in C).
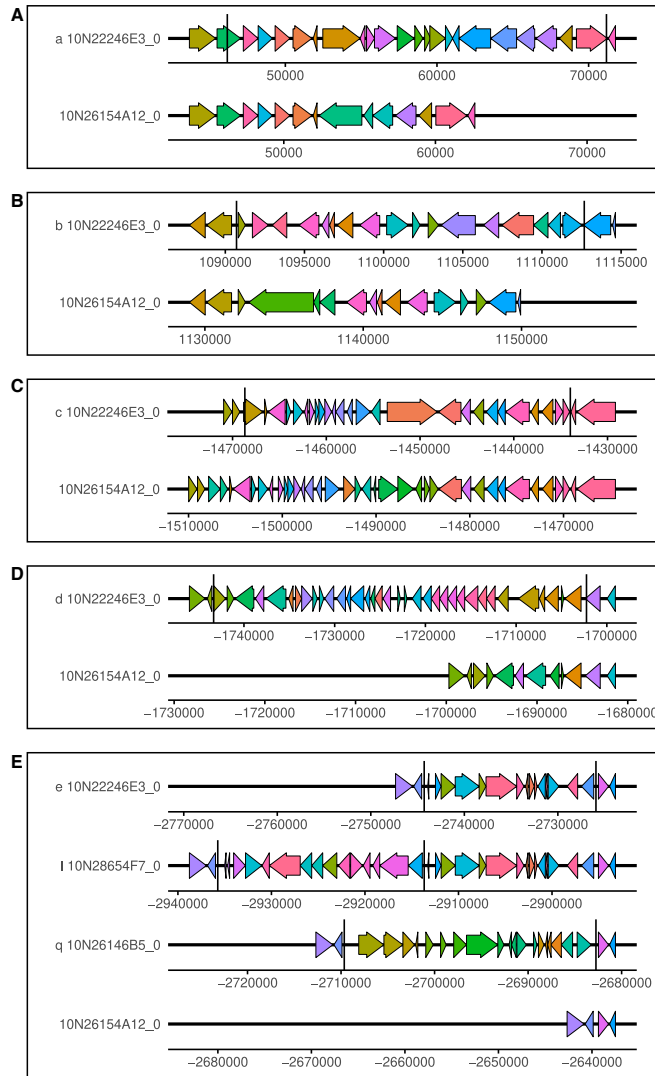
Fig. S6. (Part 1) Diagrams of all phage defense elements (PDEs) detected in the 23 *V. lentus* strains as well as examples of genomes from which PDEs are absent. PDE boundaries as determined by the k-mer-based pipeline are denoted by vertical black lines. Letters match those in Fig.1B and Data S3. Genes are colored with arbitrary colors indicating >80% sequence similarity within each figure panel. Negative genomic coordinates indicate that the region is backwards relative to the reference fasta file used to search. The double star (**) in I shows an example of diversity within the same PDE among different strains. R and S show putative PDEs that are the entirety of a small contig in the only genome where they were found and therefore the genomic context is unknown/not applicable.
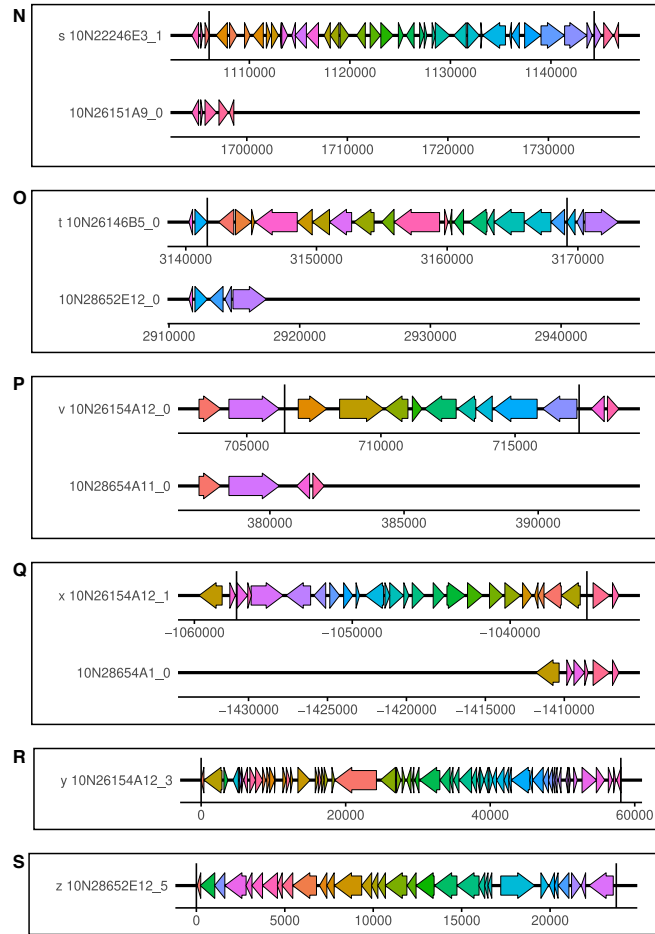
Fig. S6 (Part 2) Diagrams of all phage defense elements (PDEs) detected in the 23 *V. lentus* strains as well as examples of genomes from which PDEs are absent. PDE boundaries as determined by the k-mer-based pipeline are denoted by vertical black lines. Letters match those in Fig.1B and Data S3. Genes are colored with arbitrary colors indicating >80% sequence similarity within each figure panel. Negative genomic coordinates indicate that the region is backwards relative to the reference fasta file used to search. The double star (**) in I shows an example of diversity within the same PDE among different strains. R and S show putative PDEs that are the entirety of a small contig in the only genome where they were found and therefore the genomic context is unknown/not applicable.
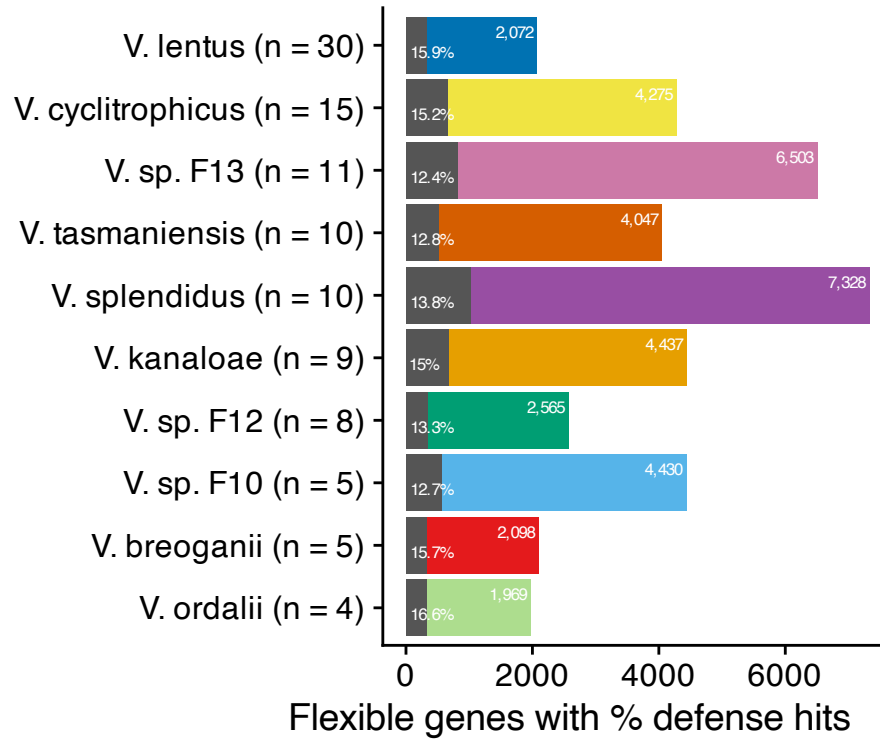
710

715

720

Fig. S6 (Part 3) Diagrams of all phage defense elements (PDEs) detected in the 23 *V. lentus*
strains as well as examples of genomes from which PDEs are absent. PDE boundaries as
725   determined by the k-mer-based pipeline are denoted by vertical black lines. Letters match those
in Fig.1B and Data S3. Genes are colored with arbitrary colors indicating >80% sequence
similarity within each figure panel. Negative genomic coordinates indicate that the region is
backwards relative to the reference fasta file used to search. The double star (**) in I shows an
example of diversity within the same PDE among different strains. R and S show putative PDEs
730   that are the entirety of a small contig in the only genome where they were found and therefore
the genomic context is unknown/not applicable.

735

Fig. S6 (Part 4) Diagrams of all phage defense elements (PDEs) detected in the 23 *V. lentus* strains as well as examples of genomes from which PDEs are absent. PDE boundaries as determined by the k-mer-based pipeline are denoted by vertical black lines. Letters match those in Fig.1B and Data S3. Genes are colored with arbitrary colors indicating >80% sequence similarity within each figure panel. Negative genomic coordinates indicate that the region is backwards relative to the reference fasta file used to search. The double star (**) in I shows an example of diversity within the same PDE among different strains. R and S show putative PDEs that are the entirety of a small contig in the only genome where they were found and therefore the genomic context is unknown/not applicable.

740

745

750

755

760

Fig. S7. Total number of flexible genes and proportion of known phage defense genes in diverse *Vibrio* species represented in different colors (n = number of genomes analyzed). Between 12-17% (grey bars) of the flexible genes of ten different species, represented as populations defined as gene flow clusters (*37*), can be attributed to known phage defense genes.

765

770
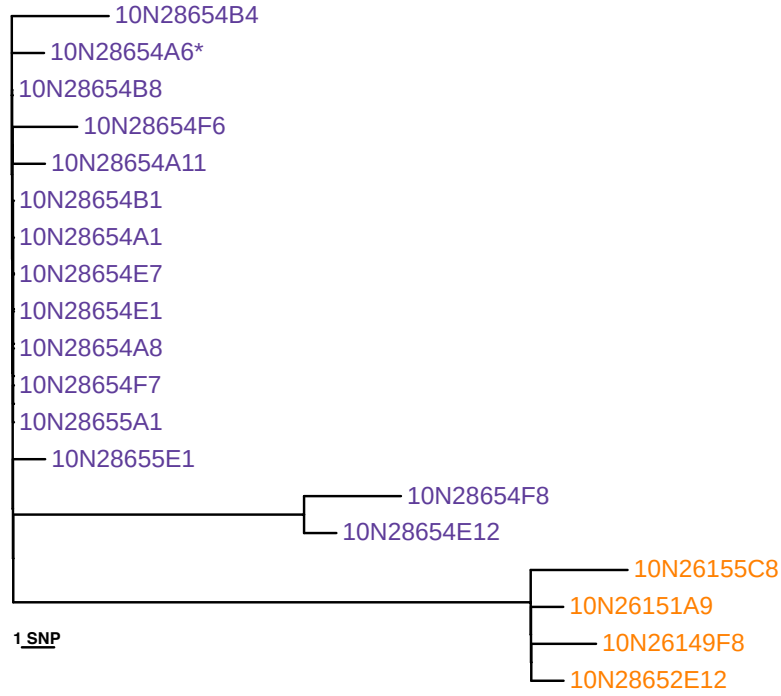
775

Fig. S8. Unrooted maximum likelihood tree for core genomes of all the nineteen "purple" and "orange" host clones. The strain chosen by the Parsnp program as a reference is indicated by *. 43 SNPs were identified in the total alignment and 14 SNPs differentiate the "orange" and "purple" subsets (see Table S3 for full list).
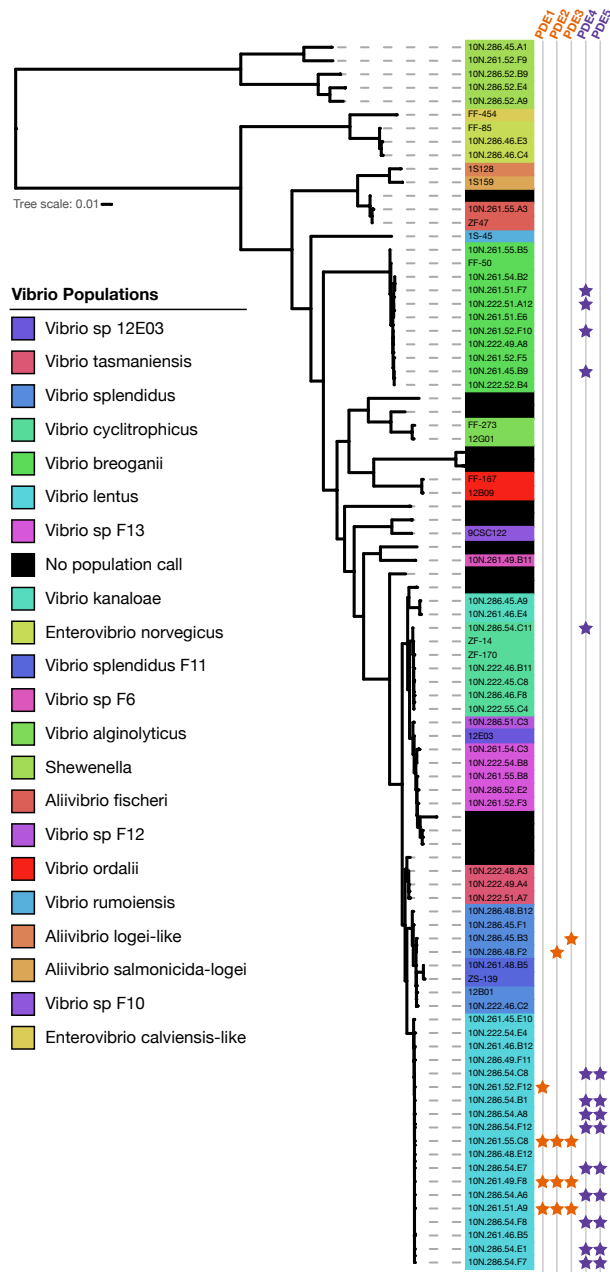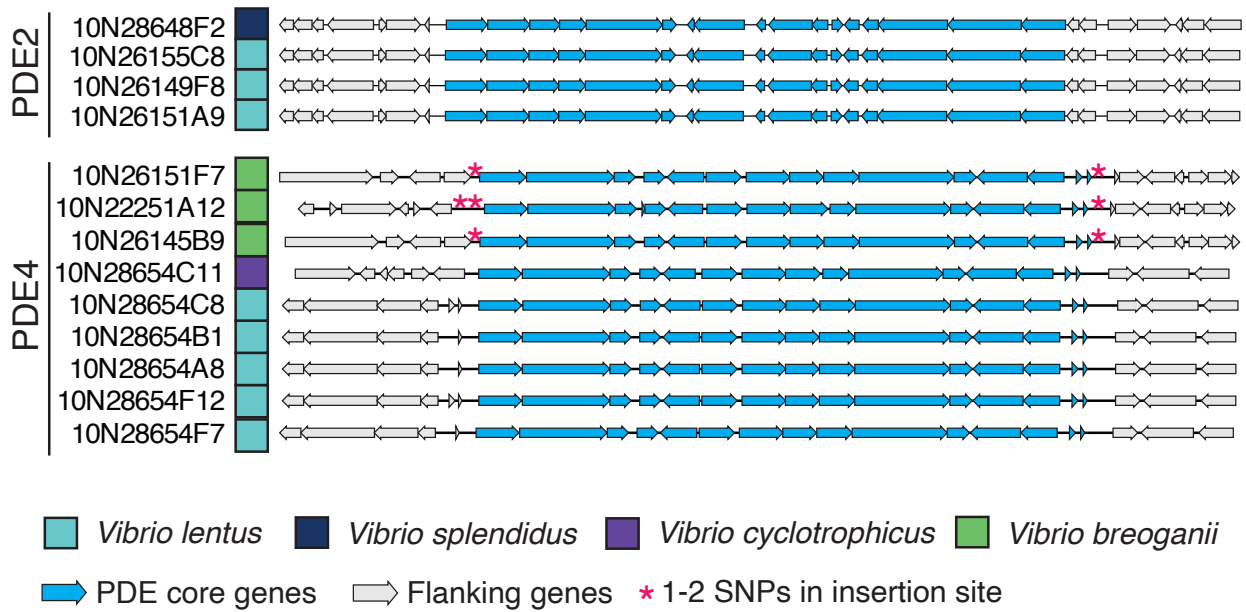
Fig. S9. Presence of the same phage defense elements (>95% nucleotide identity over >90% of the total element length) in divergent genomic backgrounds suggests their movement via horizontal gene transfer. Pruned tree from Figure S1 depicting the phylogeny of ribosomal protein and *hsp60* gene sequences (proxy for core genome) of each *Vibrio* host.

Fig. S10. Genomic context of horizontally transferred PDEs. PDEs. Specific PDEs for "orange" and "purple" phages are found in closely and distantly related *Vibrio*, sometimes in different genomic locations, but always at conserved insertion sites with the exception of a few SNPs (indicated with *). PDE2 and PDE4 in different hosts shown as examples from Fig. S9. Lower quality genomes with contig breaks at PDE insertion sites were excluded.
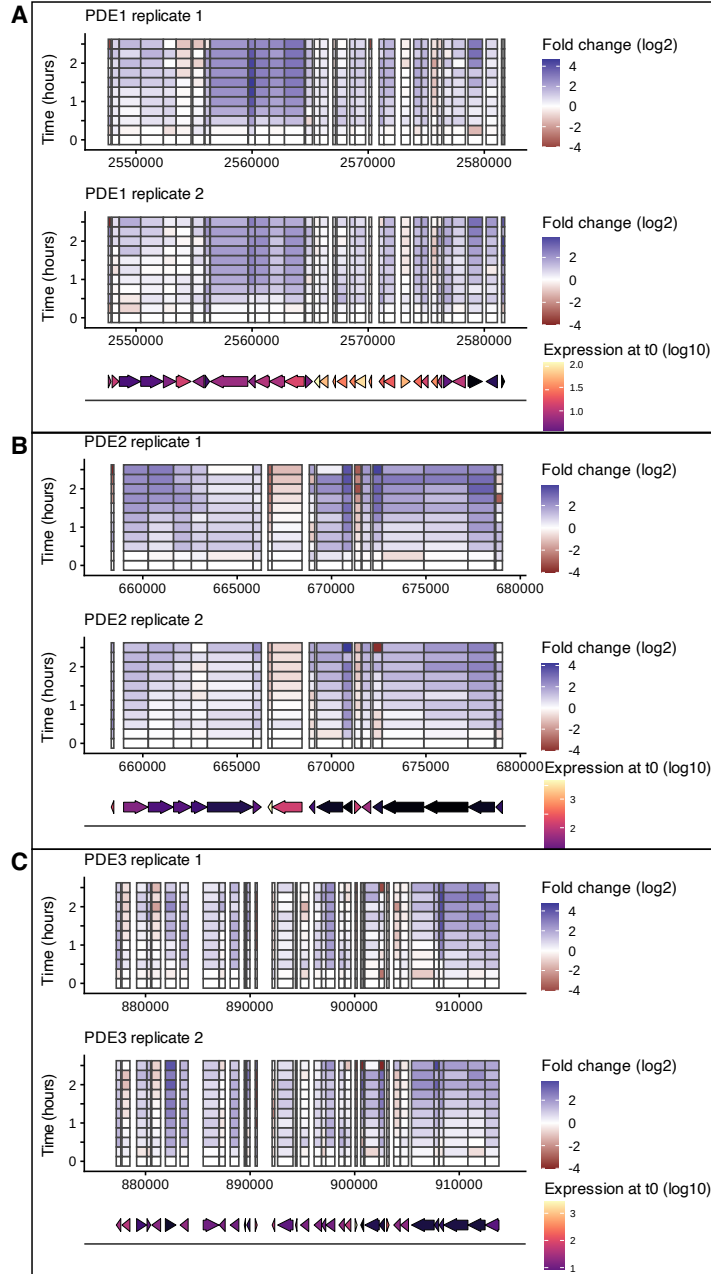
815

Fig. S11. RNA-seq experiment conducted over a time course of infection of "purple" phage 1.281.O on "orange" host 10N26155C8 shows changes in gene activity across the 3 PDEs. Boxes indicate transcribed genes and contiguous boxes indicate transcriptional units, colored by fold change in expression compared to time zero (mid exponential phase expression without phage). Gene diagrams for each PDE [(A) PDE1, (B) PDE2, (C) PDE3] are shown below the transcriptomic data summary and colored by the expression at mid exponential phase prior to phage addition (t=0). Experiment conducted with two replicates.
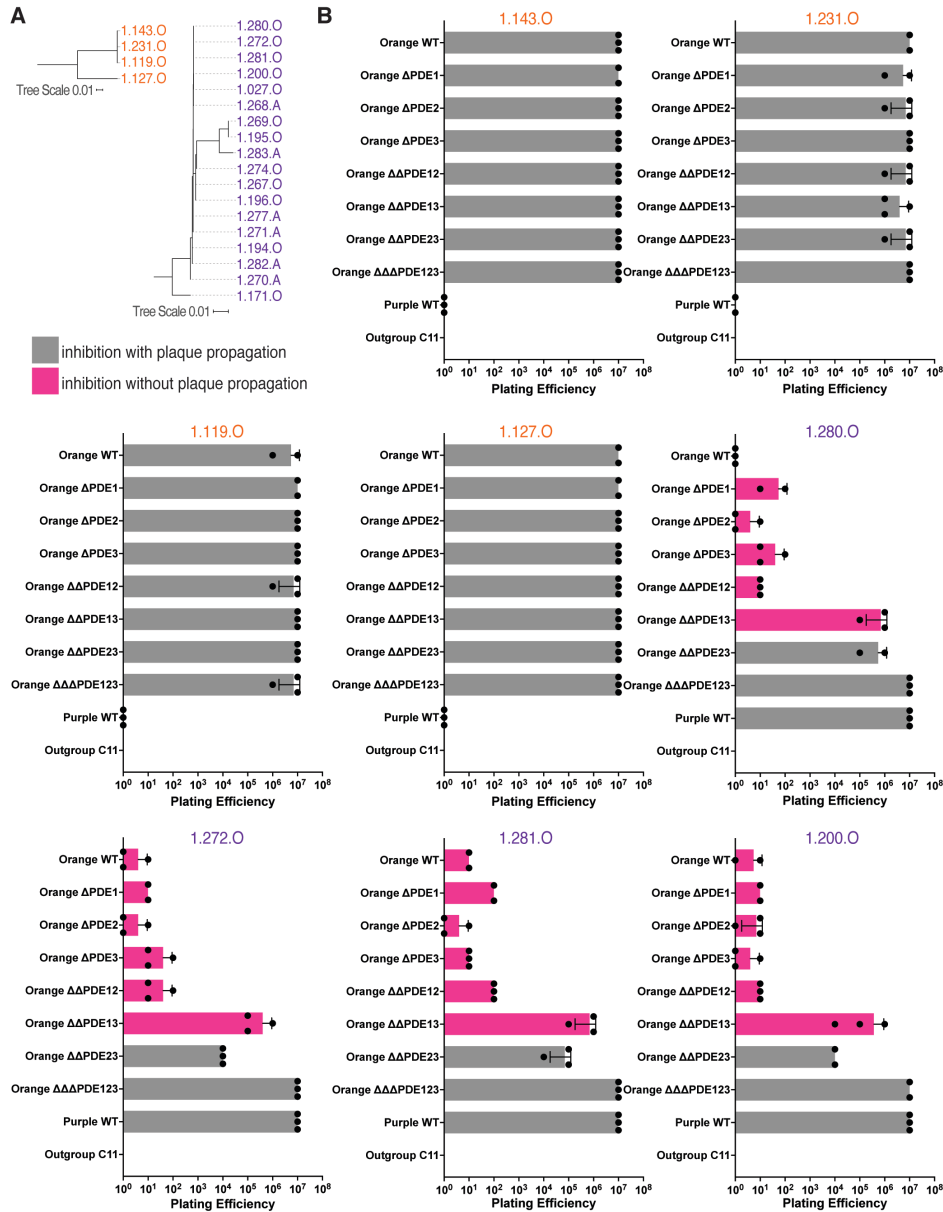
Fig. S12 (Part 1). Efficiency of plating (EOP) assay for all PDE knockouts with all "orange" and "purple" phages. (A) Phage whole genome trees (repeated from Fig. 1A). (B) EOP plating assays for each phage. Pink bars indicate inhibition without plaque propagation and grey bars indicate killing with viable phage production. Triplicates shown as data points and error bars represent standard deviation.

Fig. S12 (Part 2). Efficiency of plating (EOP) assay for all PDE knockouts with all "orange" and "purple" phages. (A) Phage whole genome trees (repeated from Fig. 1A). (B) EOP plating assays for each phage. Pink bars indicate inhibition without plaque propagation and grey bars indicate killing with viable phage production. Triplicates shown as data points and error bars represent standard deviation.
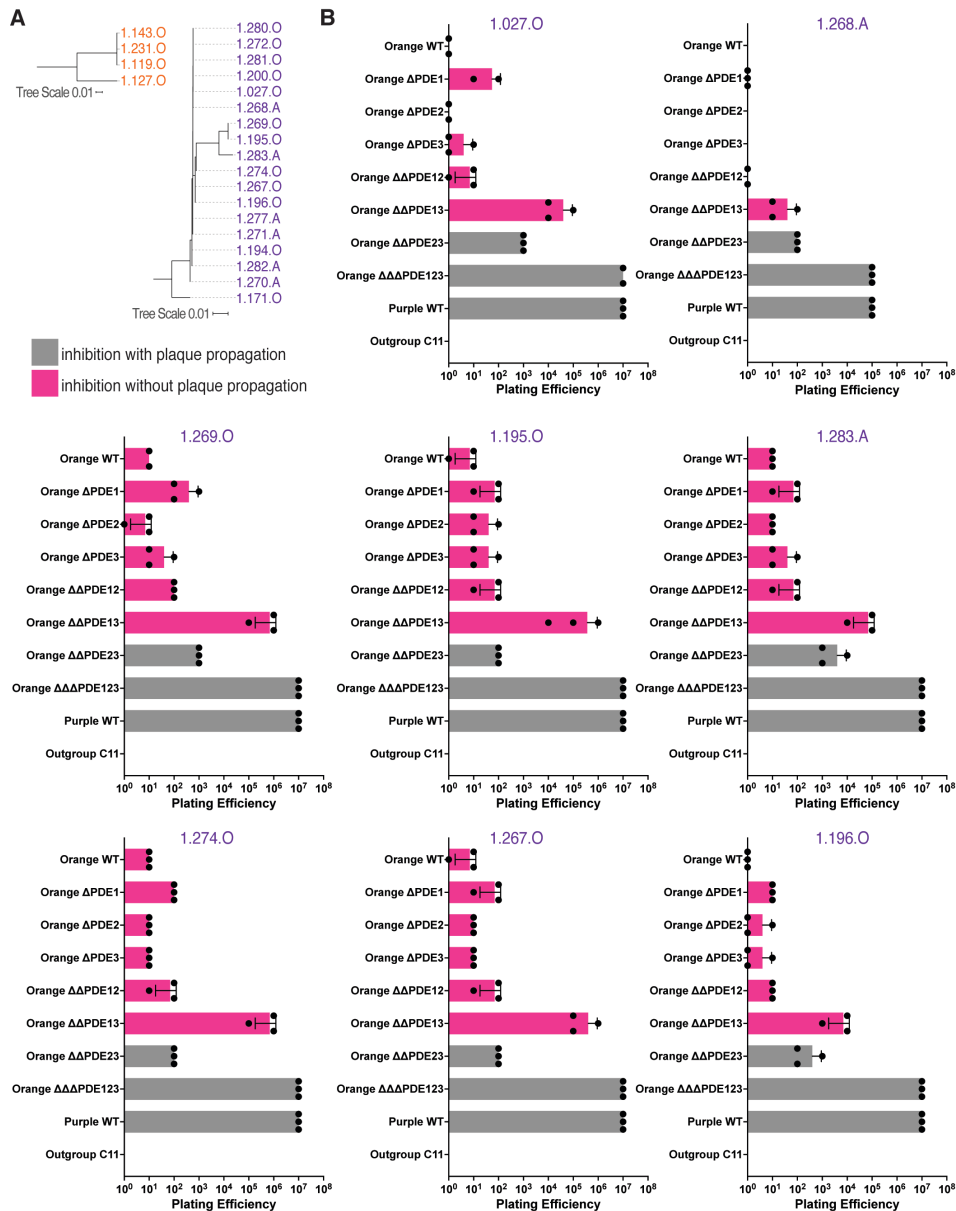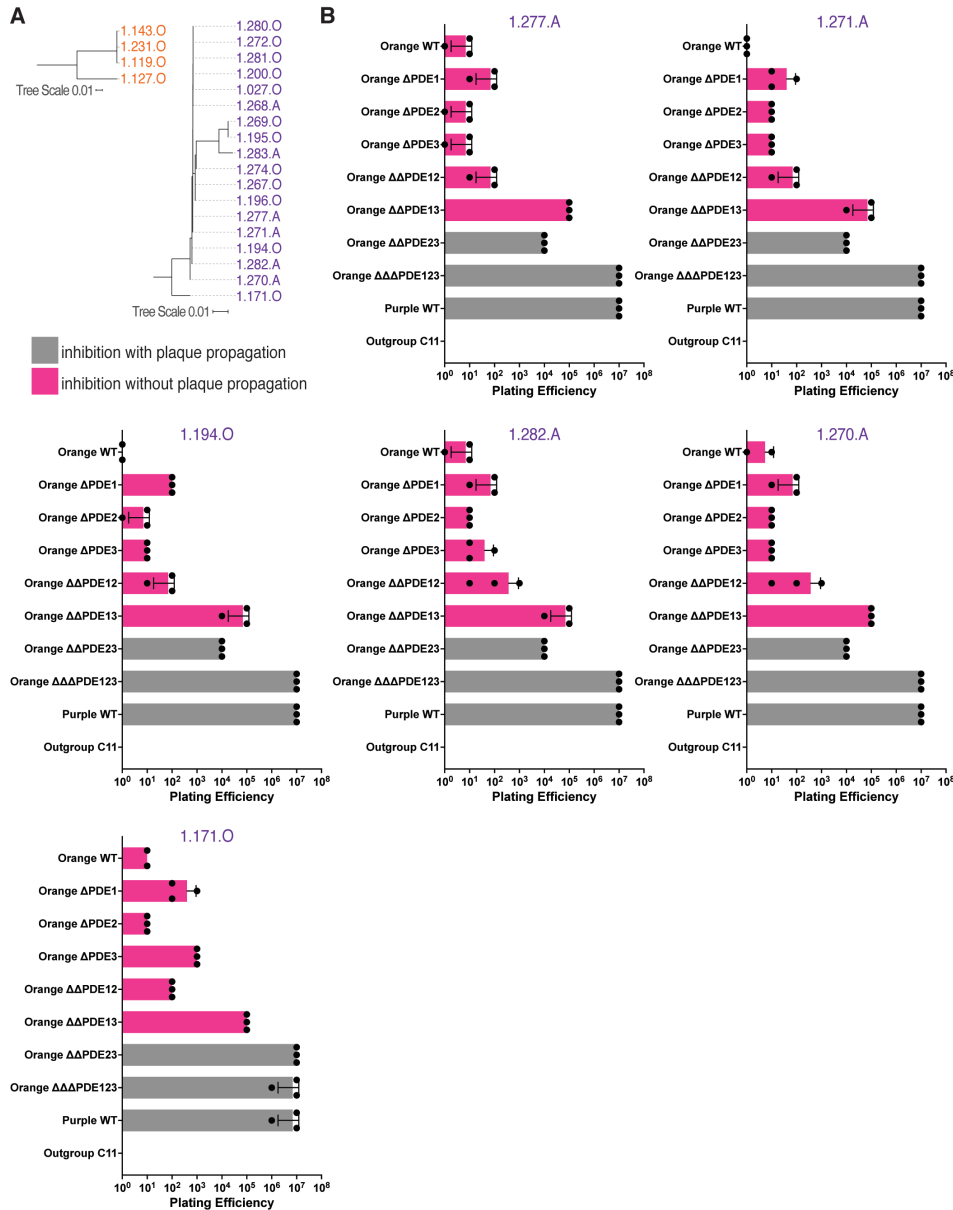
835

Fig. S12 (Part 3). Efficiency of plating (EOP) assay for all PDE knockouts with all "orange" and "purple" phages. (A) Phage whole genome trees (repeated from Fig. 1A). (B) EOP plating assays for each phage. Pink bars indicate inhibition without plaque propagation and grey bars indicate killing with viable phage production. Triplicates shown as data points and error bars represent standard deviation.
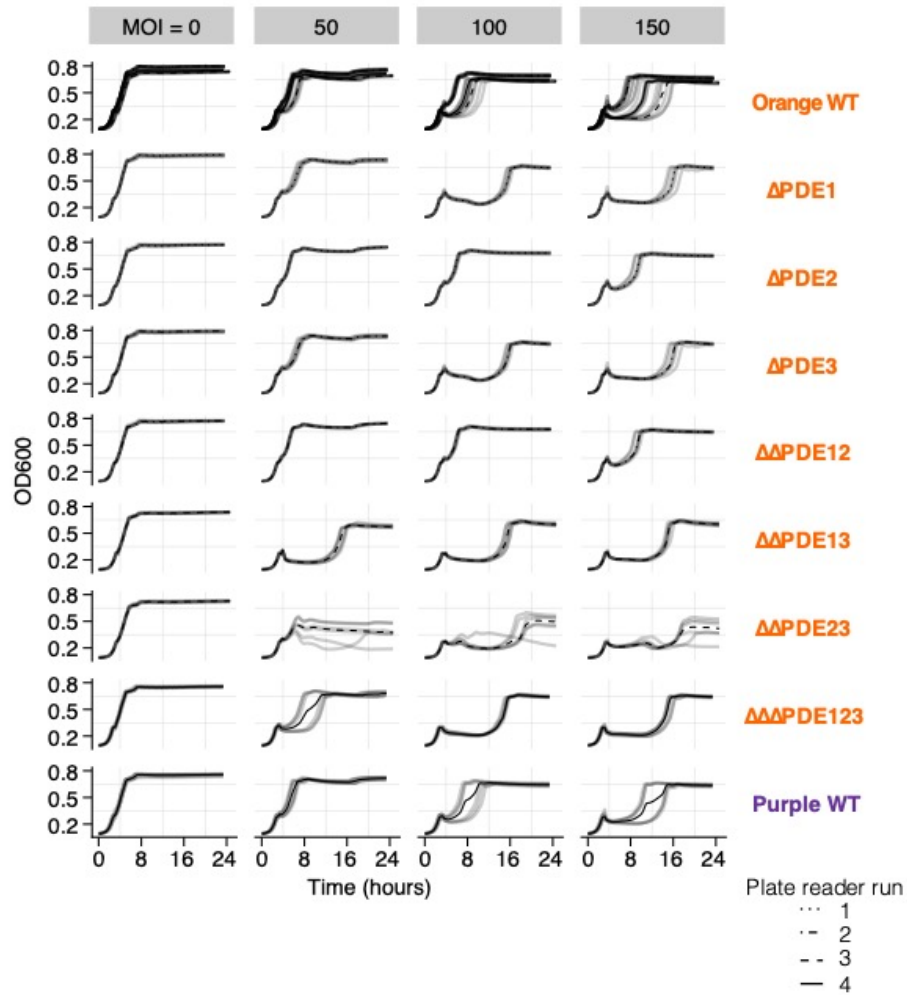
Fig. S13. PDE deletions and phage susceptibility testing in liquid. Growth curves of each combination of knockouts grown to mid-exponential phase and then challenged with "purple" phage 1.281.O at varying concentrations. Each experiment consists of three biological replicates, each done in triplicate. High MOIs were used to ensure observance of the "lysis from without" phenotype.

Fig. S14. Restriction modification phenotypic escape assayed using efficiency of plating. (A) WT "purple" phage tested on WT "purple" host compared to triple PDE knockout (left), and "purple" phage propagated on triple PDE knockout tested on both hosts (right). (B) All PDE knockouts tested against WT "purple" phage and "purple" phage propagated on triple PDE mutant. Pink bars indicate inhibition without plaque propagation and grey bars indicate killing with viable phage production. Triplicates shown as data points and error bars represent standard deviation.
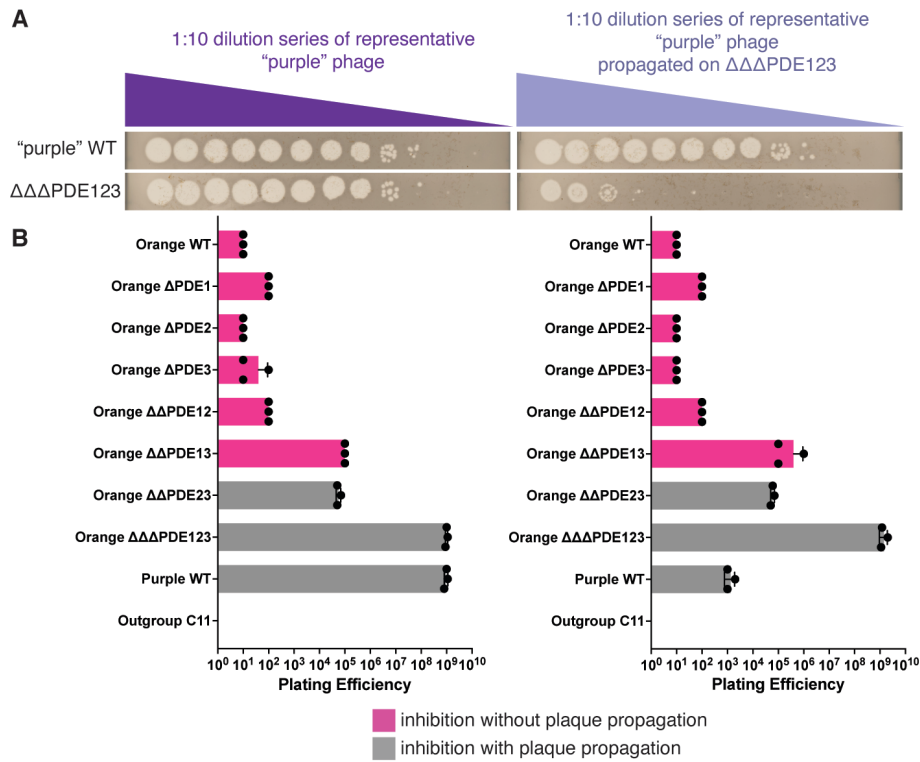
Fig. S15. Detailed genetic characterization of defense genes in PDE1. (A) Gene diagram depicting the inactivation of the restriction endonuclease (10N26155C8 ORF_0_2264) in the defense element double mutant ΔΔPDE23 using a transposon. (B) Example of efficiency of plating (EOP) assay showing that nearly full PDE deletions result in the same sensitivity to "purple" phages as inactivation of the restriction endonuclease (left). Re-streak to test for viable phage production (right). Plating done with representative "purple" phage 1.281.O. (C) Summary of replicate EOP assays. Pink bars indicate inhibition without plaque propagation and grey bars indicate killing with viable phage production. Triplicates shown as data points and error bars represent standard deviation.
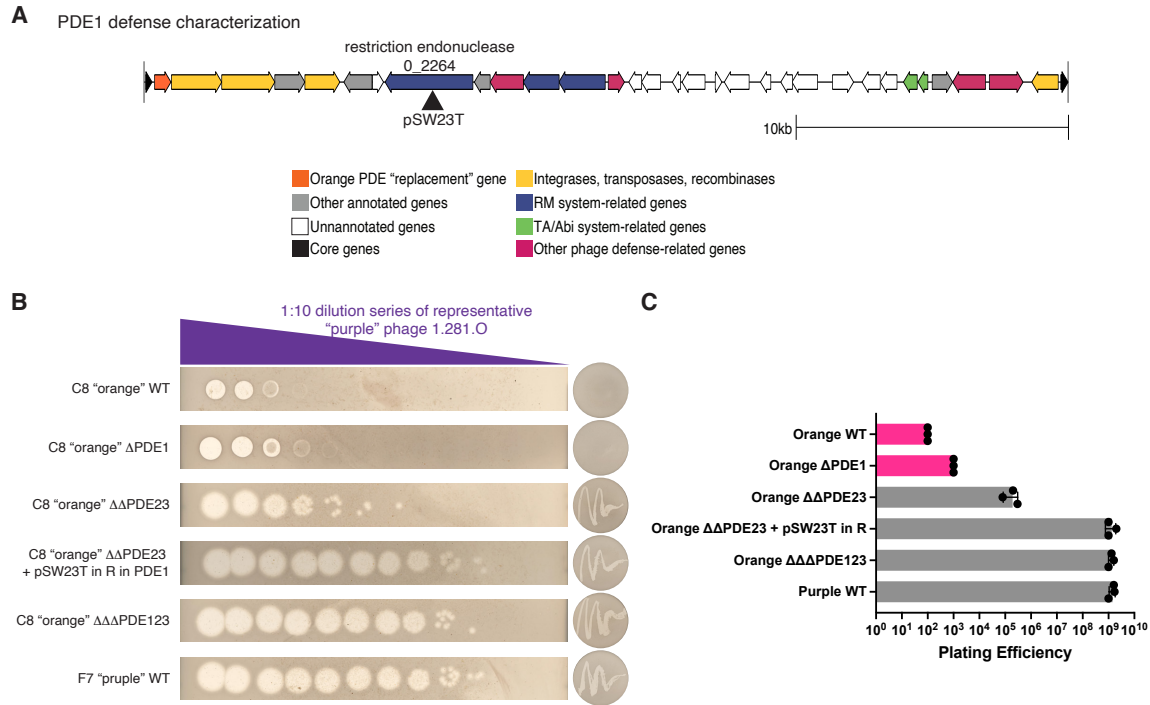
A PDE2 defense characterization

0_597: Mrr Restriction endonuclease
0_598: GTPase domain-containing protein
0_599: GTPase domain-containing protein
0_600: Unannotated
0_601: DUF4231 domain-containing protein

Legend:
- Orange PDE "replacement" gene
- Other annotated genes
- Unannotated genes
- Core genes
- Integrases, transposases, recombinases
- RM system-related genes
- TA/Abi system-related genes
- Other phage defense-related genes

B 1:10 dilution series of representative "purple" phage 1.281.O

Rows:
- C8 "orange" WT
- C8 "orange" ΔPDE2
- C8 "orange" ΔΔPDE13
- C8 "orange" ΔΔPDE13 + pSW23T in 0_597
- C8 "orange" ΔΔPDE13 + pSW23T in 0_598
- C8 "orange" ΔΔPDE13 + pSW23T in 0_599
- C8 "orange" ΔΔPDE13 + pSW23T in 0_600
- C8 "orange" ΔΔPDE13 + pSW23T in 0_601
- C8 "orange" ΔΔΔPDE123
- F7 "pruple" WT

C Plating Efficiency

Bars:
- Orange WT
- Orange ΔPDE2
- Orange ΔΔPDE13
- Orange ΔΔPDE13 + pSW23T 0_597
- Orange ΔΔPDE13 + pSW23T 0_598
- Orange ΔΔPDE13 + pSW23T 0_599
- Orange ΔΔPDE13 + pSW23T 0_600
- Orange ΔΔPDE13 + pSW23T 0_601
- Orange ΔΔΔPDE123
- Purple WT

X-axis: 10^0 10^1 10^2 10^3 10^4 10^5 10^6 10^7 10^8 10^9 10^{10} Plating Efficiency
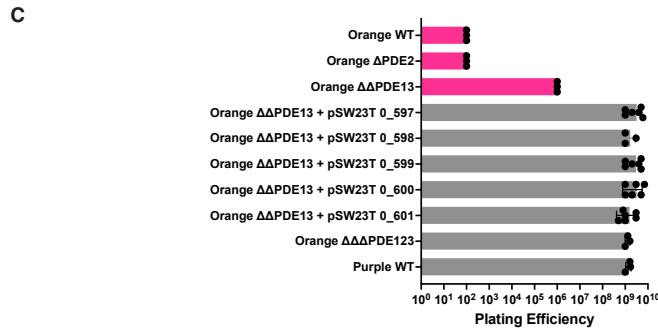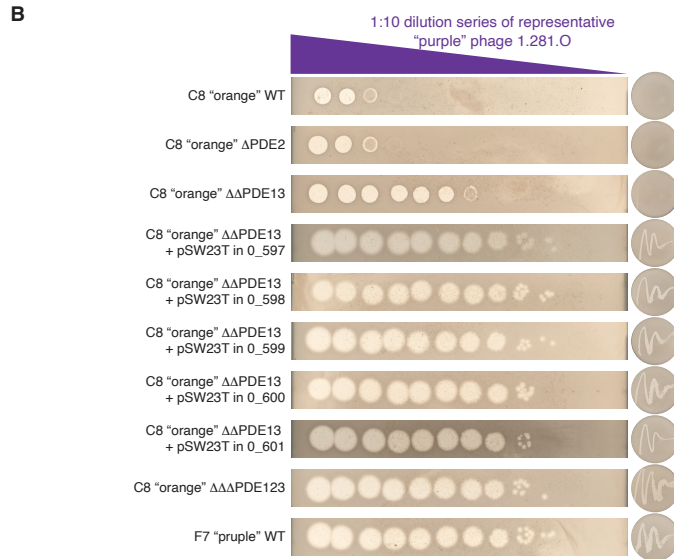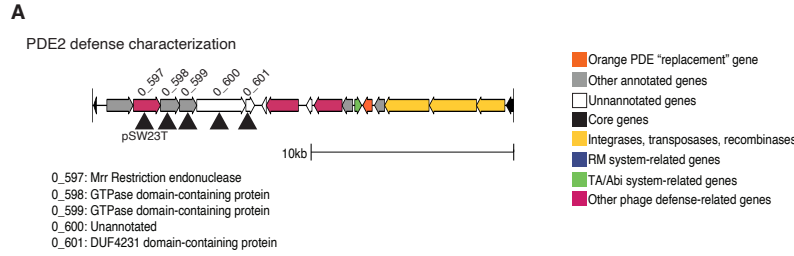
Fig. S16. Detailed genetic characterization of defense genes in PDE2. (A) Gene diagram depicting the inactivation of unannotated genes 10N26155C8 ORF_0_597, putative GTPase 10N26155C8 ORF_0_598, putative GTPase 10N26155C8 ORF_0_599, unannotated gene 10N26155C8 ORF_0_600, and unannotated gene 10N26155C8 ORF_0_601 with DUF4231 in the defense element double mutant ΔΔPDE13 using transposons. (B) Example of efficiency of plating (EOP) assay showing that nearly full PDE deletions result in the same sensitivity to "purple" phages as inactivation of each gene in A (left). Re-streak to test for viable phage production (right). Plating done with representative "purple" phage 1.281.O. (C) Summary of replicate EOP assays. Pink bars indicate inhibition without plaque propagation and grey bars indicate killing with viable phage production. Triplicates shown as data points and error bars represent standard deviation.
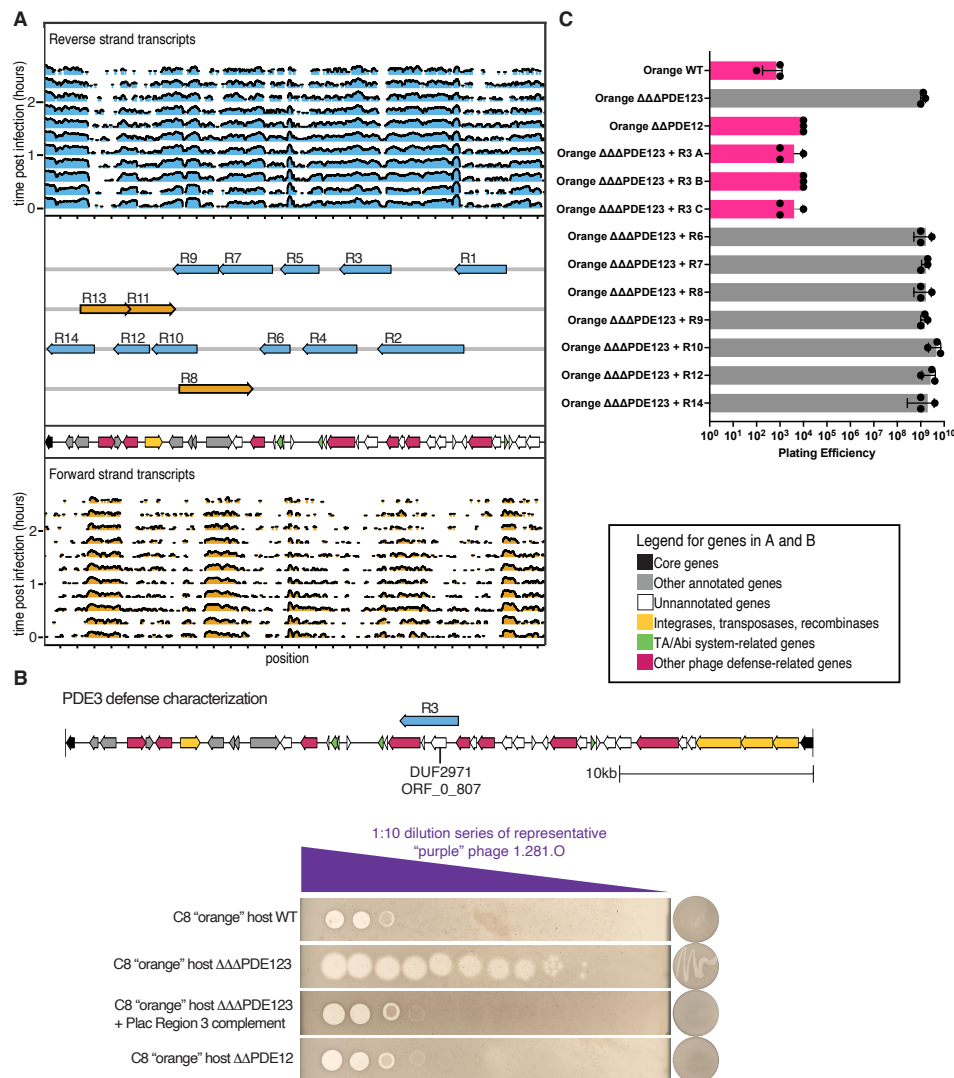
Fig. S17. Detailed genetic characterization of defense genes in PDE3. (A) Identification of co-transcribed regions (R1-R14) during a time course of infection of "orange" host 10N26155C8 with "purple" phage 1.281.O. (B, top) Gene diagram depicting complementation of R3 in the triple defense element mutant ΔΔΔPDE123. R3 being the only region to restore resistance phenotype. (B, bottom) Example of efficiency of plating (EOP) assay showing that the presence of PDE3 results in the same resistance to "purple" phages as the complementation of R3 in ΔΔΔPDE123 (bottom, left). Re-streak to test for viable phage production (bottom, right). Plating done with representative "purple" phage 1.281.O. (C) EOP assay for all regions complemented in to triple PDE knockout. Pink bars indicate inhibition without plaque propagation and grey bars indicate killing with viable phage production. Triplicates shown as data points and error bars represent standard deviation.
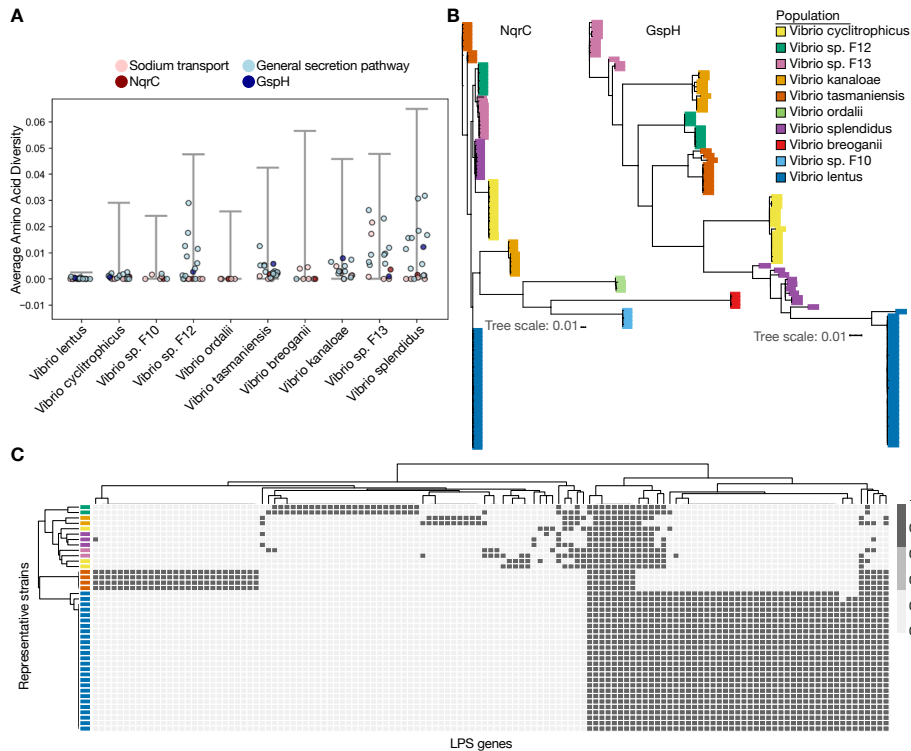
Fig. S18. Phage receptor diversity across diverse *Vibrio* species suggests positive selection at the population level. (A) Genes identified as receptors in *V. lentus* show much reduced diversity compared to core genes in 10 populations of vibrios. Populations are defined as co-existing members of species and represent gene flow units (*37*). Each point represents the average amino acid pairwise diversity for a specific receptor gene indicated by color. The grey bars represent the 2.5th and 97.5th percentile range of average amino acid pairwise diversity for all core genes within a given population. (B) Nucleotide trees of select receptor genes, sodium transporter (*nqrC*) and pseudopilin (*gspH*), show evidence of recent population-level sweeps since receptor alleles within each population are nearly identical. (C) Genes encoding the LPS show the same presence/absence patterns within populations indicating that they encode the same LPS variant, indicating receptor evolution is congruent with population evolution. Populations are defined here as gene flow clusters that also represent ecological units (*37*). The recognition of population boundaries is key for interpretation of gene or allele frequency in the context of differential selective forces.

935

940

945

950

955
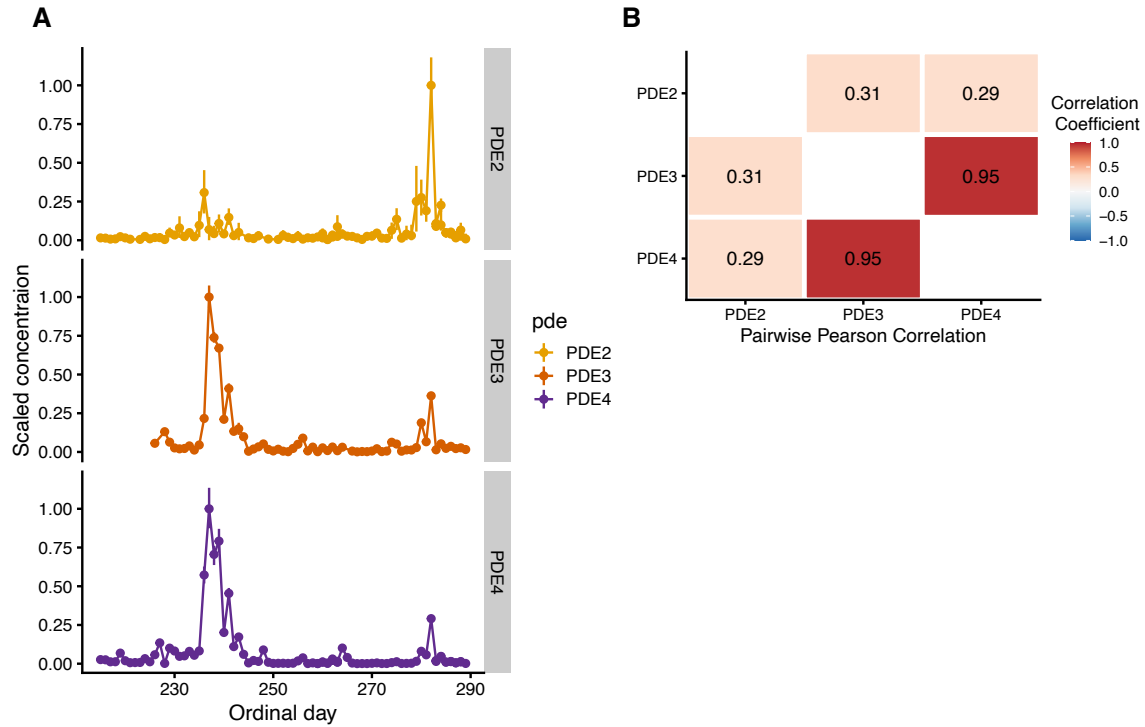
Fig. S19.Droplet digital PCR quantification of PDEs over time in the Nahant time series. (A) Relative abundance of PDE2 and PDE3 in "orange" clonal genomic background and PDE4 in "purple" clonal genomic background were quantified by designing an amplicon that spans the insertion site of each PDE in its specific genomic background. Error bars represent Poisson 95% confidence intervals. (B) Pairwise Pearson correlation of PDE relative abundances over the time series.

970

Fig. S20. Distribution of putative phage defense regions in closely related *Listeria* strains. (A) Core genome tree with presence/absence matrix of elements. (B) Flexible genome percentage breakdown of phage defense elements. (C) Accompanying gene diagrams of each phage defense element. In gene diagrams, hits to known defense genes are shown in green. See supplemental Data S4 for genes and annotations.
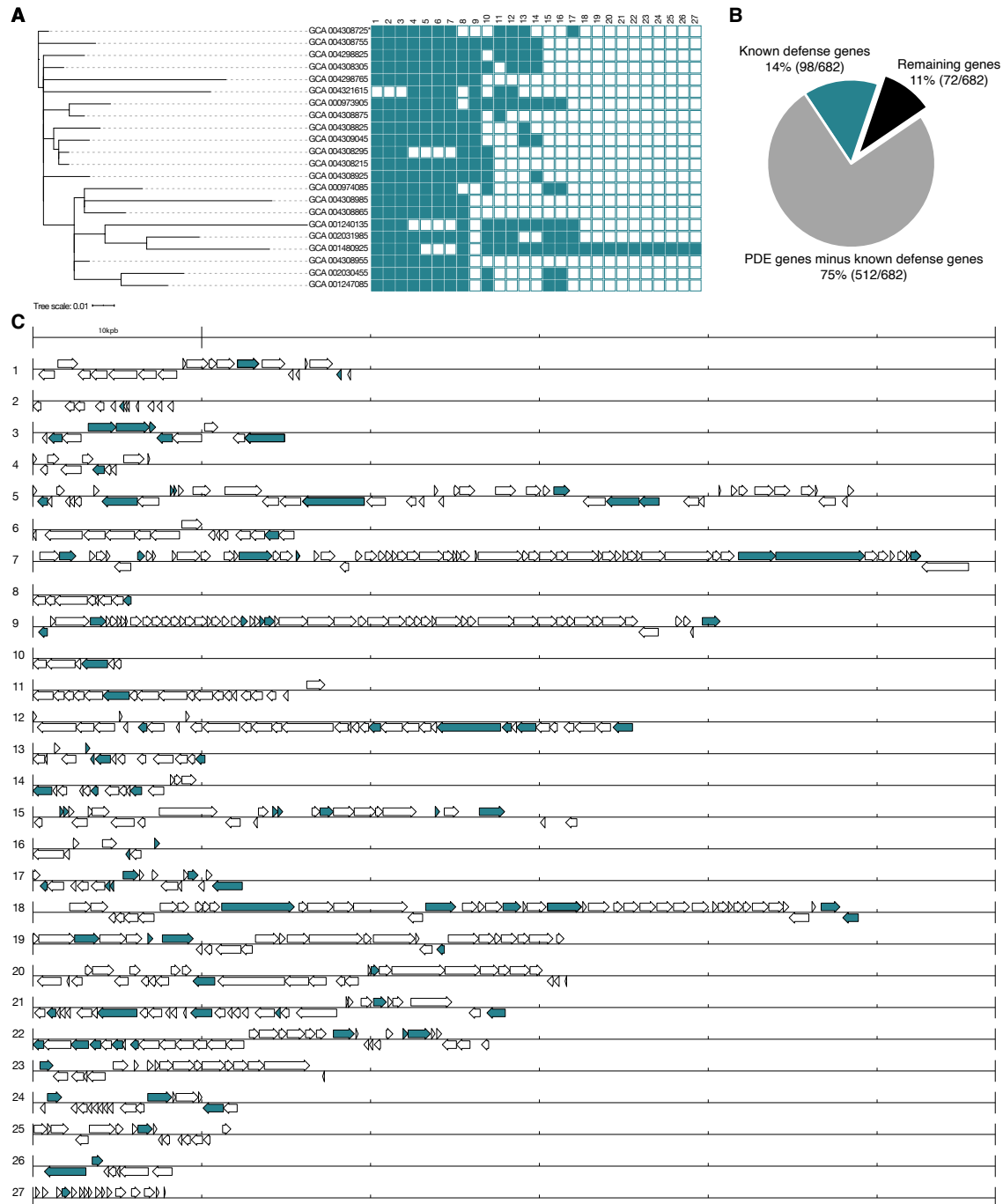
Fig. S21.Distribution of putative phage defense regions in closely related *Salmonella* strains. (A) Core genome tree with presence/absence matrix of elements. (B) Flexible genome percentage breakdown of phage defense elements. (C) Accompanying gene diagrams of each phage defense element. In gene diagrams, hits to known defense genes are shown in blue. See supplemental Data S5 for genes and annotations.
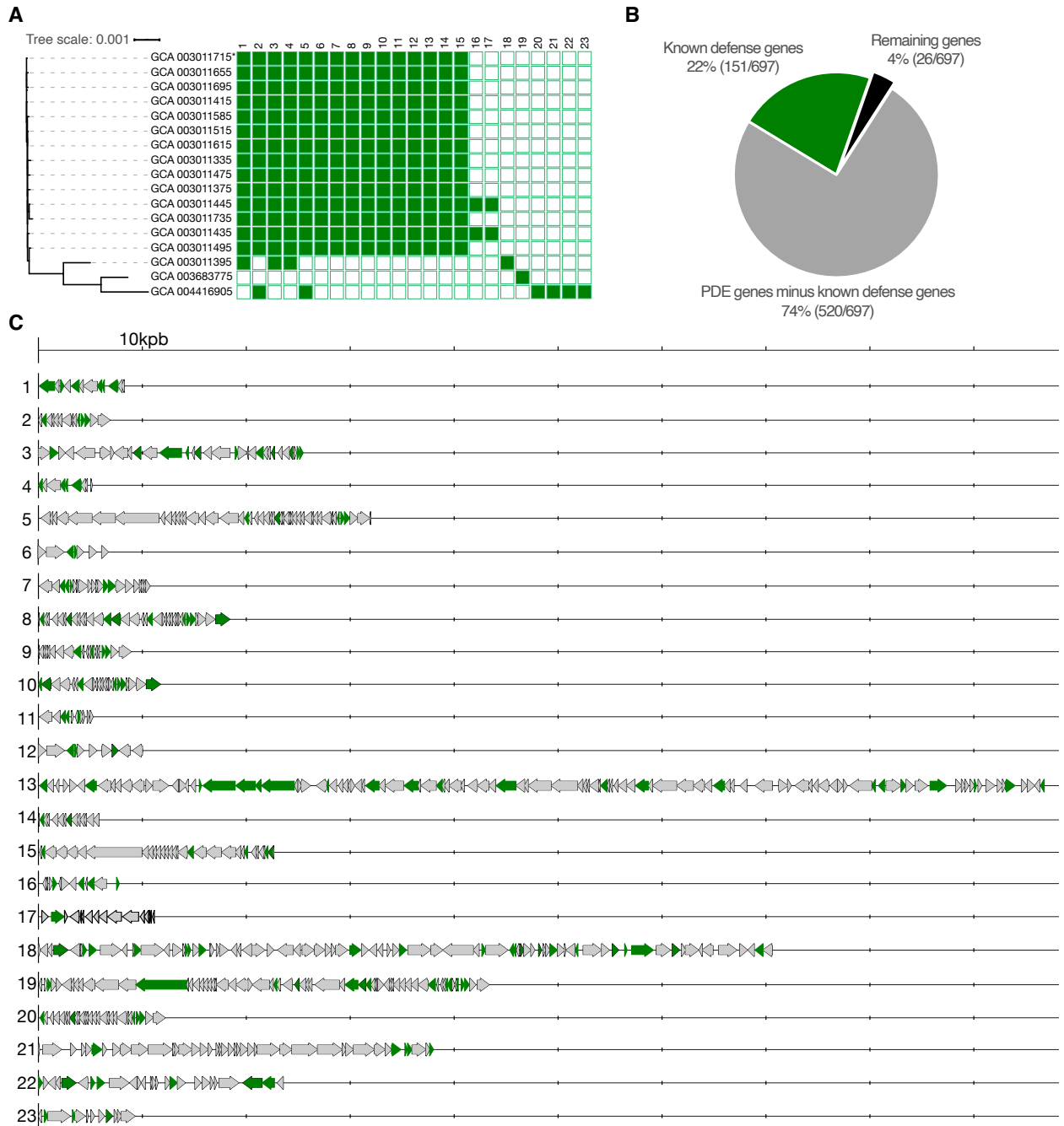
Fig. S22.Distribution of putative phage defense regions in closely related *Clostridium* strains. (A) Core genome tree with presence/absence matrix of elements. (B) Flexible genome percentage breakdown of phage defense elements. (C) Accompanying gene diagrams of each phage defense element. In gene diagrams, hits to known defense genes are shown in red. See supplemental Data S6 for genes and annotations.
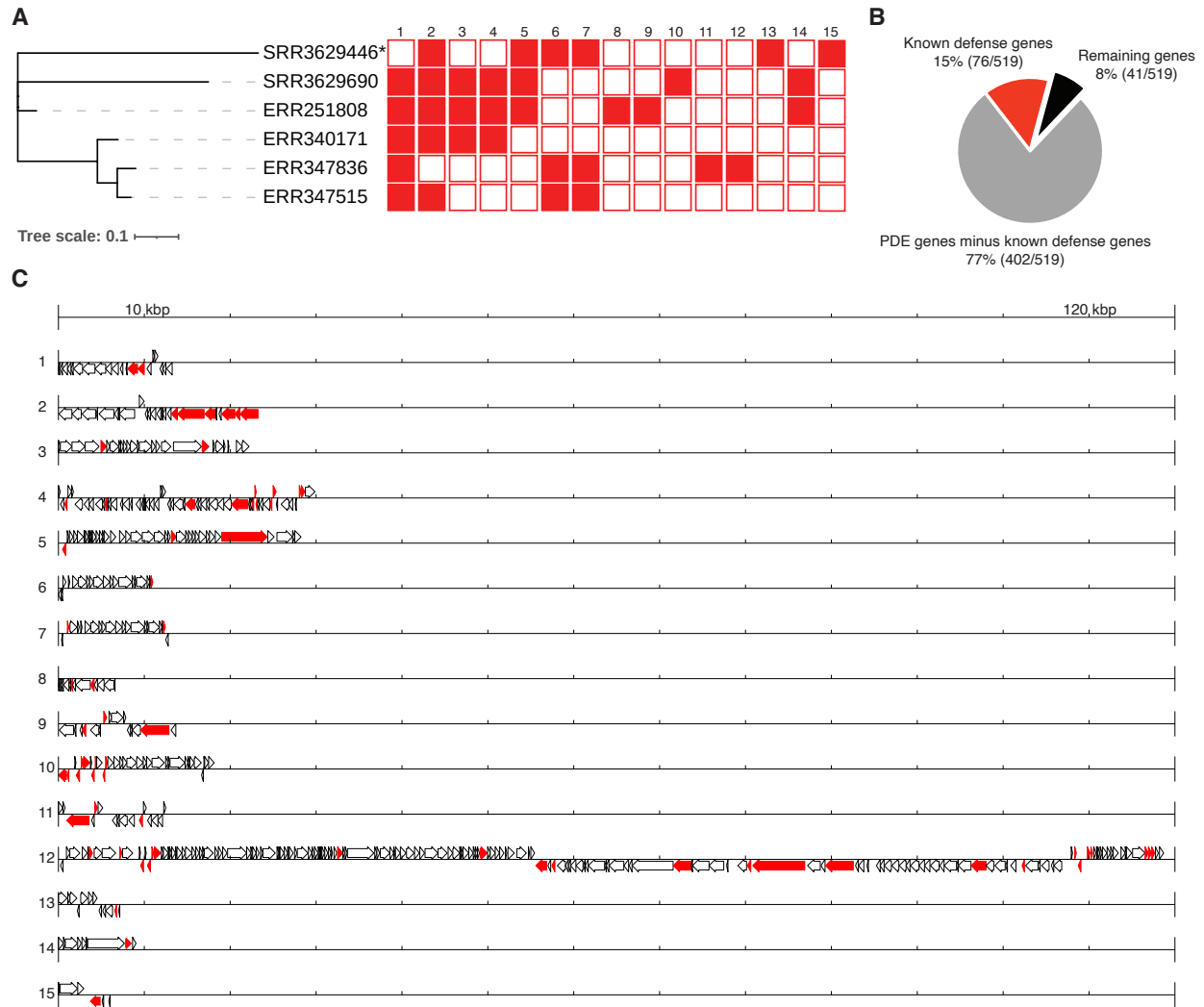
995   Table S1.

Receptor identification by transposon mutagenesis. All mappings used 10N26155C8 as a reference.

Phage	Gene ID	Annotation	Tn Hits
1.143.O	10N26155C8_2_159	General secretion pathway protein H (*gspH*)	81
1.143.O	10N26155C8_2_164	General secretion pathway protein C (*gspC*)	9
1.143.O	10N26155C8_2_162	General secretion pathway protein E (*gspE*)	1
1.143.O	10N26155C8_2_156	General secretion pathway protein K (*gspK*)	1
1.281.O	10N26155C8_2_94	dTDP-4-dehydrorhamnose reductase (lps3)	58
1.281.O	10N26155C8_2_93	dTDP-4-dehydrorhamnose 3,5-epimerase	8
1.281.O	10N26155C8_2_96	dTDP-glucose 4,6-dehydratase	6
1.281.O	10N26155C8_0_669	Phosphoglucomutase	3
1.281.O	10N26155C8_0_485	ABC-type multidrug transport system, ATPase and permease component	1
1.281.O	10N26155C8_0_2482	Na(+)-translocating NADH-quinone reductase subunit B (nqr*B*)	1
1.281.O	10N26155C8_0_2480	Na(+)-translocating NADH-quinone reductase subunit D (nqrD)	1

1000

Table S2.

Receptor identification by sequencing of spontaneous resistant mutants. Single nucleotide variants (SNVs) and deletions were identified using CLC work bench 9 workflow or though further manual identification using coverage analysis (CA). All mappings used 10N26155C8 as a reference.

Phage	Replicate	Contig_Position	Type	Length	Gene	AA change	Annotation
1.119.O	1	2_171942	Deletion	1	ORF_2_160	ORF_2_160:p.Lys7fs	General secretion pathway protein G (*gspG*)
1.119.O	2	2_173977	SNV	1	ORF_2_162	ORF_2_162:p.Lys267*	General secretion pathway protein E (*gspE*)
1.119.O	3	2_1671-167305	Deletion via CA	109	ORF_2_154	NA	General secretion pathway protein M (*gspM*)
1.119.O	4	2_167254	SNV	1	ORF_2_154	ORF_2_154:p.Trp63*	General secretion pathway protein M (*gspM*)
1.119.O	5	2_173221	Deletion	1	ORF_2_161	ORF_2_161:p.Gly18fs	General secretion pathway protein F (*gspF*)
1.119.O	6	2_1719-172018	Deletion via CA	83	ORF_2_160	NA	General secretion pathway protein G (*gspG*)
1.119.O	6	2_166050-166174	Deletion via CA	125	ORF_2_153	NA	General secretion pathway protein N (*gspN*)
1.281.O	1	2_2804974	SNV	1	ORF_0_2483	ORF_0_2483:p.Ser302*	Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-) (n*qrA*)
1.281.O	2	2_68299	Insertion	1	ORF_2_61	ORF_2_61:p.Tyr187fs	Unannotated
1.281.O	3	2_72111-72115	Deletion via CA	5	NA	NA	intergenic, upstream of 10N26155C8_2_65 (Unannotated)
1.281.O	3	2_72128-72134	Deletion via CA	7	NA	NA	intergenic, upstream of 10N26155C8_2_65 (Unannotated)
1.281.O	4	2_68299	Insertion	1	ORF_2_61	ORF_2_61:p.Tyr187fs	Unannotated
1.281.O	5	2_108233	SNV	1	ORF_2_96	ORF_2_96:p.Gln332*	dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)
1.281.O	6	2_68845-68906	Deletion via CA	62	ORF_2_61&62	NA	end of 10N26155C8_2_61 and beginning of 10N26155C8_2_62
1.281.O	7	2_107805	SNV	1	ORF_2_95	ORF_2_95:p.Asn109Tyr	Glucose-1-phosphate thymidylyltransferase (EC 2.7.7.24)

Table S3.
Core genome SNPs separating "purple" and "orange" clones (matching Fig. S8). Strain
10N28654F7 was used as the reference.

1010

SNP #	Nucleotide change P-->O	ORF in 10N28654F7	Annotation	AA change P-->O	Notes
1	G-->T	ORF_0_1115	Cytochrome O ubiquinol oxidase subunit III	Val-->Phe	
2	C-->T	ORF_0_1604	Unannotated	NA	Asp-->Asp; Near end of a PDE
3	C-->T	ORF_1_1456	Pesticidial crystal protein cry6Aa	Gln-->Stop	
4	G-->A	ORF_1_917	Aerobic cobaltochelatase CobS subunit	Ala-->Val	
5	G-->A	ORF_1_912	TRAP-type C4-dicarboxylate transport system, periplasmic component	NA	Gly-->Gly
6	A-->G	ORF_1_365-366	Unannotated	Stop-->Gln	
7	C-->T	ORF_0_3133	Phosphoenolpyruvate carboxylase	NA	Gly-->Gly
8	C-->T	ORF_0_1978	IcmF-related protein	Ala-->Val	
9	A-->G	ORF_1_87	Unannotated	Thr-->Ala	
10	A-->G	intergenic	NA	NA	upstream of ORF_1_86: putative orphan protein; putative membrane protein
11	C-->G	ORF_1_687	RND multidrug efflux transporter; Acriflavin resistance protein	Gln-->Glu	
12	G-->A	ORF_0_1847	Formate dehydrogenase -O, gamma subunit	Met-->Ile	
13	C-->A	intergenic	NA	NA	upstream of ORF_0_138:tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA
14	G-->A	ORF_1_1585	Probable MFS transporter	NA	Ala-->Ala

Table S4.
Predicted motifs of RM systems on PDEs and corresponding methylome data.

RM PDE	PDE 1	
Predicted Motif	TCA*BN(4)RTRTC	
	Host/Phage	Fraction Methylated
	10N26155C8	599/600
	1.119.O	1/1
	1.127.O	1/1
	1.143.O	1/1
	1.231.O	1/1
	10N28654F7	0/599
	1.283.A	0/7
	1.281.O	0/8
	1.196.O	0/8
RM PDE	PDE 4	
Predicted Motif	CCA*GN(6)TAA	
	Host/Phage	Fraction Methylated
	10N26155C8	0/646
	1.119.O	0/3
	1.127.O	0/4
	1.143.O	0/3
	1.231.O	0/3
	10N28654F7	635/638
	1.283.A	5/5
	1.281.O	4/4
	1.196.O	4/4
RM PDE	PDE 5	
Predicted Motif	GA*GN(6)GGC	
	Host/Phage	Fraction Methylated
	10N26155C8	0/1795
	1.119.O	0/17
	1.127.O	0/15
	1.143.O	0/17
	1.231.O	0/17
	10N28654F7	1770/1779
	1.283.A	8/8
	1.281.O	10/10
	1.196.O	10/10