MICROBIOLOGY

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

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]*, Martin F. Polz[1,3]*

Although it is generally accepted that phages drive bacterial evolution, how these dynamics play out in the wild remains poorly understood. We found 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 to 12 defense elements, accounting for more than 90% of the flexible genome among close relatives. The rapid turnover of these elements decouples phage resistance from other genomic features. Thus, resistance to phages in the wild follows evolutionary trajectories alternative to those predicted from laboratory-based evolutionary experiments.

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 one reason for the 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 coevolution 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 because such modifications frequently incur fitness costs (*6*). Indeed, many additional defense mechanisms have recently been discovered, including CRISPR systems (*14*) and several other innate immunity mechanisms, many of which remain to be mechanistically characterized (*15*–*19*). Genes encoding both receptors and defense systems have been shown to occur frequently in variable genomic islands (*6*, *20*–*22*) or to be associated with mobile genetic elements (*23*–*26*). For CRISPR-based adaptive immunity, which is present in more than 40% of bacteria (*27*), evolutionary processes have been documented on ecological time scales (*28*, *29*). By contrast, 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 the evolution of wild microbial populations.

Here, we combined population genomic and molecular genetic approaches to determine 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 more than 1300 strains of marine *Vibrio* ranging in relatedness from near-clonal to species-level divergence (*34*, *37*). *Vibrio* hosts isolated on three different days (ordinal dates 222, 261, and 286) were used as "bait" to isolate 248 co-occurring lytic viruses by quantitative plaque assays, and each of 245 plaque-positive hosts was 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 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 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; see also fig. S2, A to D, and data S1). Testing all pairwise interactions of these phages with all potential hosts at varying multiplicity of infection (MOI) revealed 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 found that although the two phage types use different receptors, all bacterial genomes encode both sets of receptors. Molecular genetic analysis suggests that the "orange" phage receptor is the type II secretion system (T2SS) pseudopilus (GspH), whereas the "purple" phages have two receptors: LPS, an extracellular receptor, and the sodium transporter NqrC, a membrane receptor (*39*) (fig. S5 and 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 that receptor mutations play a secondary role, the data overall support the idea 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), revealed 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

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Department of Microbiology and Parasitology, University of Santiago de Compostela, 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, Vienna, Austria. [4]Ifremer, Unité Physiologie Fonctionnelle des Organismes Marins, 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 author. Email: martin.f.polz@univie.ac.at (M.F.P.); fleroux@sb-roscoff.fr (F.L.R.)
†These authors contributed equally to this work.

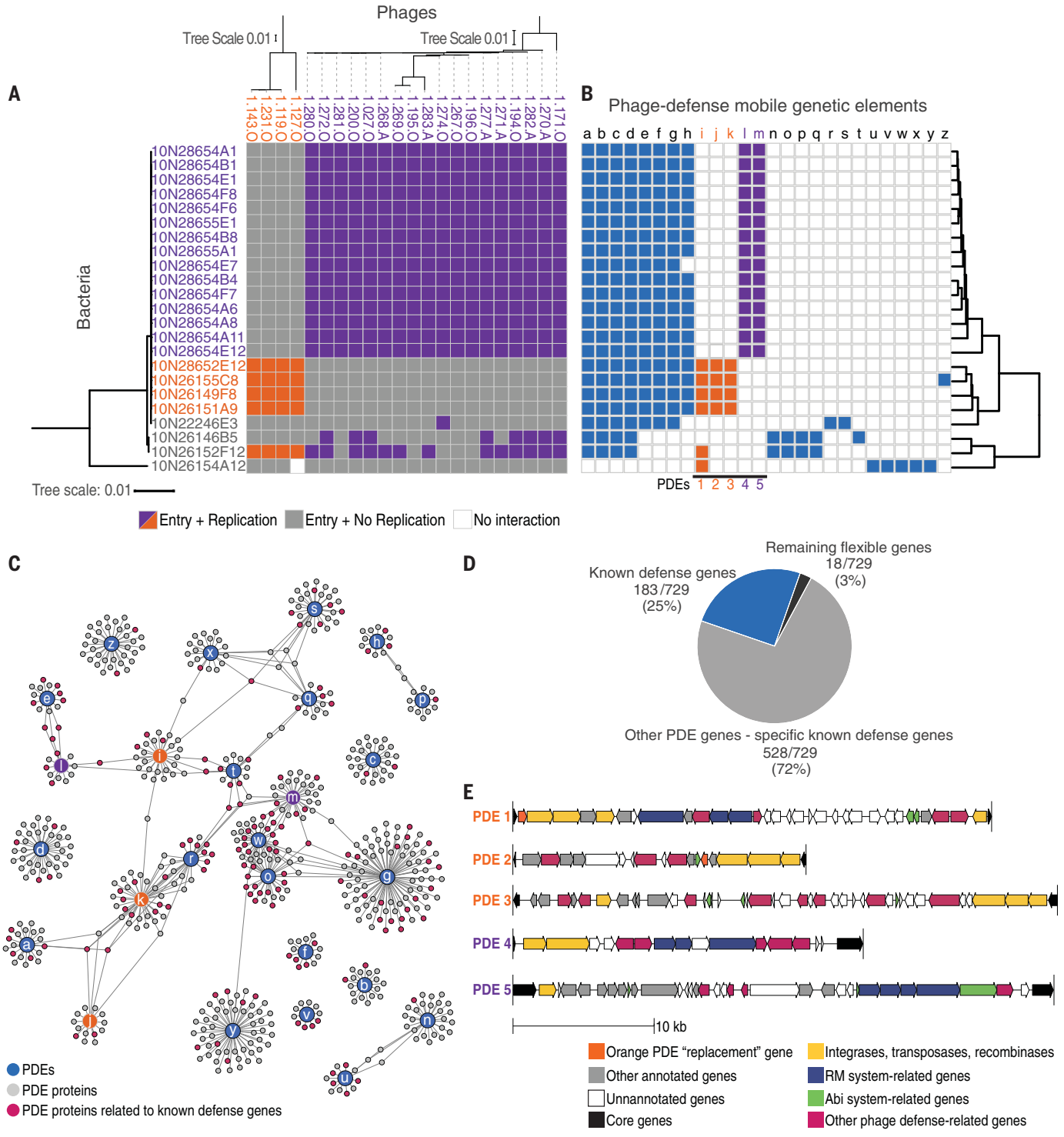

Fig. 1. Near-clonal strains of *Vibrio lentus* differ in sensitivity to phage predation and in their carriage of mobile genetic elements encoding for phage defense genes. (**A**) Phage host range matrix. Rows represent bacterial strains; columns represent phages. Left: Core genome phylogenetic tree of hosts. Top: Whole-genome tree of viruses. (**B**) Distribution of 26 PDEs in 23 hosts. PDEs responsible for defense of the "orange" and "purple" host groups are indicated in the same colors. Right: Hierarchical clustering of presence or 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. PDEs are indicated in blue, and the orange and purple PDEs are color-coded accordingly. Gray nodes are genes; defense genes are indicated in magenta. (**D**) Fraction of the bacterial flexible genome attributed to phage defense. Among the 23 clones, an all-by-all genomic comparison shows that 97% of flexible regions greater than 5 kbp are putative PDEs. Only 25% of the total flexible genes match known defense genes; 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.

pairwise genome comparisons (data S2), we identified a total of 28 flexible genome regions (>5 kb 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 8500 to 89,600 bp and contain genes annotated as recombinases/integrases and transposases (data S3), but the elements appear overall unrelated because only a few genes are shared (Fig. 1C and 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) because 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 more than 95% of the flexible genomic regions (Fig. 1D and 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 observed a similar range of 12 to 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 that 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 differentiated by only 14 SNPs across the core genome, yet are differentially infected by the two phage groups (Fig. 1A, fig. S8, and table S3). The flexible genome of these "orange" and "purple" hosts is differentiated by three (PDEs 1 to 3) and two (PDEs 4 and 5) large PDEs, respectively (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 guanosine triphosphatases (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 that 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 investigated the extent to which each of the PDEs contributes to the differential resistance observed between the "purple" and "orange" hosts. All three PDE-encoded RM systems (PDEs 1, 4, and 5; Fig. 1E) appear active according to 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 and figs. S12 and 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 because deletion of all three PDEs is required for full sensitivity.

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 by a factor of 10 (from 10^{-1} to 10^{-2} phage dilution) but did not yield viable phage progeny when assayed using a restreaking test (Fig. 2C). This phenotype led us to hypothesize that there may be a multilevel defense structure involving the remaining two PDEs. Knocking out PDE2 and PDE3 independently resulted in no change from the wild-type 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 RM escape mutants (*39*) (fig. S14). Knocking out PDE1 and PDE2 together resulted in the same factor-of-10 increase in killing observed when deleting PDE1 alone, whereas knocking out PDE1 and PDE3 together yielded cell death at much lower MOIs (10^{-5} phage dilution); however, 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 "purple" wild-type host strains. Therefore, we conclude that all three elements are needed for full, wild-type–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. Although we were only able to find a single gene (the restriction enzyme) necessary for defense on PDE1 (Fig. 2A and fig. S15), three unannotated genes and two putative GTPases within a module on the 5′ end of PDE2 appear involved in defense (Fig. 2A and fig. S16). For PDE3, we inactivated 11 different genes as well as two regions associated with 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 cotranscribed 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 (Fig. 2A and fig. S17, B and C). Hence, two of the three PDEs carry uncharacterized phage defense genes.

The invariance of receptor genes 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 that 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 are 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,

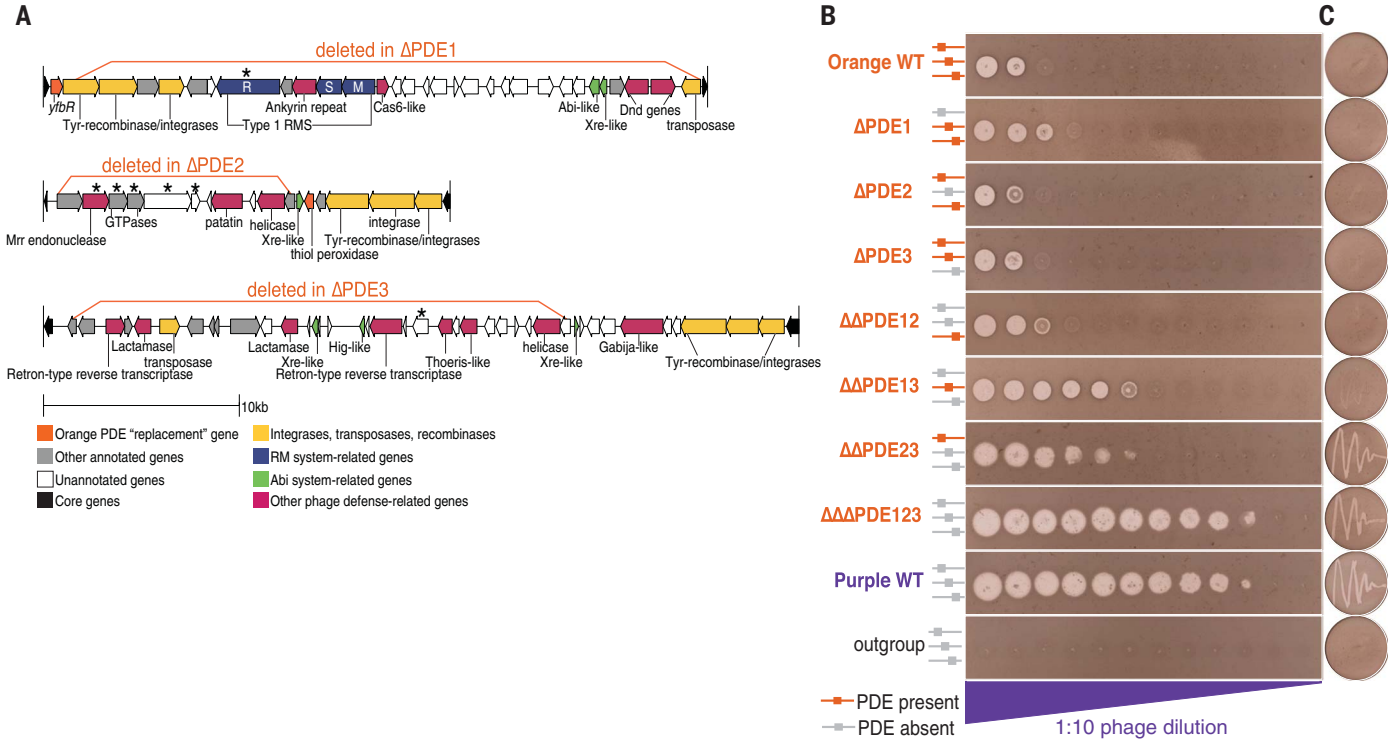

Fig. 2. Changes in susceptibility to phage killing observed for phage defense element (PDE) deletions. (**A**) Detailed gene diagrams indicate markerless deletions of defense gene–containing portions of the different "orange" PDEs as well as genes determined to be responsible for defense phenotypes (marked with an asterisk; see figs. S15 to S17). (**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**) Restreak 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 restreak plates.

which frequently serves as the primary receptor for many phages, also appears similar at the population level because the genes responsible for its synthesis display population-specific presence/absence patterns suggesting that 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 its existence indicates a key difference from the lab, where, in rich media, receptor mutations frequently arise in phage-host cocultures (*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 be advantageous only under extreme predation regimes, or under regimes of low effective diversity, such as a clonal infection or laboratory culture.

If defense is largely determined by differential PDE carriage at the clonal level, this 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 because genomes with zero SNPs can differ by at least one putative PDE, and the "purple" and "orange" host 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 observed a differential shift in relative PDE abundance. We designed droplet digital polymerase chain reaction (ddPCR) 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). PDE3 and PDE4 (representing "orange" and "purple" genetic backgrounds, respectively) were near perfectly correlated across the entire time series, whereas PDE2 gained in importance only toward 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 *Vibrio* species—follow a primarily nonrecombinogenic mode of evolution (*37*, *46*–*48*), we have observed 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 and data S4), *Salmonella* (fig. S21 and data S5), and *Clostridium* (fig. S22 and 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 coexisting diversity in many natural bacterial populations is high (*49*), resistance at the clonal level effectively lowers prey concentration and thus the frequency of 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, although 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 because 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. It is therefore necessary to determine the mechanisms of transfer and the life history strategies of PDEs. 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, *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000).
2. M. Breitbart, F. Rohwer, *Trends Microbiol.* **13**, 278–284 (2005).
3. H. G. Hampton, B. N. J. Watson, P. C. Fineran, *Nature* **577**, 327–336 (2020).
4. K. E. Kortright, B. K. Chan, J. L. Koff, P. E. Turner, *Cell Host Microbe* **25**, 219–232 (2019).
5. S. J. Labrie, J. E. Samson, S. Moineau, *Nat. Rev. Microbiol.* **8**, 317–327 (2010).
6. S. Avrani, O. Wurtzel, I. Sharon, R. Sorek, D. Lindell, *Nature* **474**, 604–608 (2011).
7. J. R. Meyer *et al*., *Science* **335**, 428–432 (2012).
8. W. Arber, *Annu. Rev. Microbiol.* **19**, 365–378 (1965).
9. G. G. Wilson, N. E. Murray, *Annu. Rev. Genet.* **25**, 585–627 (1991).
10. G. Bertani, J. J. Weigle, *J. Bacteriol.* **65**, 113–121 (1953).
11. I. J. Molineux, *New Biol.* **3**, 230–236 (1991).
12. L. Snyder, *Mol. Microbiol.* **15**, 415–420 (1995).
13. A. Lopatina, N. Tal, R. Sorek, *Annu. Rev. Virol.* **7**, 371–384 (2020).
14. R. Barrangou *et al*., *Science* **315**, 1709–1712 (2007).
15. S. Doron *et al*., *Science* **359**, eaar4120 (2018).
16. A. Millman *et al*., *Cell* **183**, 1551–1561.e12 (2020).
17. L. Gao *et al*., *Science* **369**, 1077–1084 (2020).
18. R. P. Novick, G. E. Christie, J. R. Penadés, *Nat. Rev. Microbiol.* **8**, 541–551 (2010).
19. A. A. Govande, B. Duncan-Lowey, J. B. Eaglesham, A. T. Whiteley, P. J. Kranzusch, *Cell Rep.* **35**, 109206 (2021).
20. F. Rodriguez-Valera *et al*., *Nat. Rev. Microbiol.* **7**, 828–836 (2009).
21. K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, *J. Bacteriol.* **193**, 6039–6056 (2011).
22. A. Bernheim, R. Sorek, *Nat. Rev. Microbiol.* **18**, 113–119 (2020).
23. N. D. McDonald, A. Regmi, D. P. Morreale, J. D. Borowski, E. F. Boyd, *BMC Genomics* **20**, 105 (2019).
24. A. C. McKitterick, K. D. Seed, *Nat. Commun.* **9**, 2348 (2018).
25. E. V. Koonin, K. S. Makarova, Y. I. Wolf, M. Krupovic, *Nat. Rev. Genet.* **21**, 119–131 (2020).
26. K. N. LeGault *et al*., *Science* **373**, eabg2166 (2021).
27. K. S. Makarova *et al*., *Nat. Rev. Microbiol.* **9**, 467–477 (2011).
28. A. F. Andersson, J. F. Banfield, *Science* **320**, 1047–1050 (2008).
29. E. Laanto, V. Hoikkala, J. Ravantti, L.-R. Sundberg, *Nat. Commun.* **8**, 111 (2017).
30. T. Dimitriu, M. D. Szczelkun, E. R. Westra, *Curr. Biol.* **30**, R1189–R1202 (2020).
31. A. Chevallereau, B. J. Pons, S. van Houte, E. R. Westra, *Nat. Rev. Microbiol.* 10.1038/s41579-021-00602-y (2021).
32. S. van Houte, A. Buckling, E. R. Westra, *Microbiol. Mol. Biol. Rev.* **80**, 745–763 (2016).
33. K. M. Kauffman *et al*., *Sci. Data* **5**, 180114 (2018).
34. K. M. Kauffman *et al*., *Nature* **554**, 118–122 (2018).
35. K. M. Kauffman *et al*., BioRxiv 449121 [preprint]. 27 June 2021.
36. A. M. Martin-Platero *et al*., *Nat. Commun.* **9**, 266 (2018).
37. P. Arevalo, D. VanInsberghe, J. Elsherbini, J. Gore, M. F. Polz, *Cell* **178**, 820–834.e14 (2019).
38. M. Delbrück, *J. Gen. Physiol.* **23**, 643–660 (1940).
39. See supplementary materials.
40. O. X. Cordero, M. F. Polz, *Nat. Rev. Microbiol.* **12**, 263–273 (2014).
41. B. J. Shapiro *et al*., *Science* **336**, 48–51 (2012).
42. E. R. Westra *et al*., *Curr. Biol.* **25**, 1043–1049 (2015).
43. S. Zborowsky, D. Lindell, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 16899–16908 (2019).
44. E. O. Alseth *et al*., *Nature* **574**, 549–552 (2019).
45. C. A. Hernandez, B. Koskella, *Evolution* **73**, 2461–2475 (2019).
46. K. E. Dingle *et al*., *PLOS ONE* **6**, e19993 (2011).
47. B. J. Shapiro, *Curr. Opin. Microbiol.* **31**, 116–123 (2016).
48. Y. Yin *et al*., *Microbiol. Res.* **175**, 84–92 (2015).
49. J. R. Thompson *et al*., *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 are provided in data S9 and online (DOI: 10.5061/dryad.sj3tx965k).

ACKNOWLEDGMENTS

We thank M. Cutler for experimental support; D. Newman, K. Costa, and H. Wildschutte for advice on mutagenesis experiments; O. X. Cordero, S. Kearney, and A. F. Takemura for valuable suggestions throughout; S. W. Chisholm, L. Kelly, and S. Pollak for thoughtful comments on the manuscript; and D. Kwon, X. Yu, O. Abudayyeh, and J. Gootenberg for use of ddPCR equipment and/or lab space during the SARS-CoV-2 pandemic. **Funding:** Simons Foundation (Life Sciences Project Award-572792) (M.F.P.); National Science Foundation Division of Ocean Sciences (OCE-1435868) (M.F.P.); MIT J-WAFS seed grant (M.F.P.); NSF GRFP (F.A.H.); MIT Martin Society of Fellows for Sustainability (F.A.H.); Xunta de Galicia Postdoctoral Fellowship (ED481B 2016/032) (J.D.); Agence Nationale de la Recherche (ANR-16-CE32-0008-01) (F.L.R.); European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 884988, Advanced ERC Dynamic) (F.L.R.). **Author contributions:** Conceptualization: F.A.H., M.F.P. Data curation: F.A.H., with contributions from J.E., J.D., M.M., P.A., D.V., B.K.R.-J., K.K. Methodology: F.A.H., J.D., M.M., J.E., F.L.R. Investigation: F.A.H., J.D., F.L.R., with contributions from J.E., M.M., P.A., D.V., K.K., B.K.R.-J., H.G. Validation: F.A.H., J.D., F.L.R., with contributions from J.E., H.G., A.L., A.G. Visualization: F.A.H., J.E., with contributions from D.V., P.A., K.K. Software: F.A.H., J.E., M.M., with contributions from D.V., P.A. Funding acquisition: M.F.P., F.L.R., F.A.H., J.D. Project administration and supervision: M.F.P., F.L.R. Writing–original draft: F.A.H., M.F.P. Writing–review and editing: F.A.H., M.F.P., with contributions from J.D., F.L.R. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** New genomes used in this work have been deposited under the NCBI BioProject with accession number PRJNA328102. Main code is provided in data S2 and online (*50*). Supporting data are provided in data S9 and online (*51*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abb1083
Materials and Methods
Figs. S1 to S22
Tables S1 to S4
References (*52*–*97*)
Data S1 to S9

1 February 2020; resubmitted 10 May 2021
Accepted 1 September 2021
10.1126/science.abb1083