



# Balance between promiscuity and specificity in phage $\lambda$ host range

Bryan Andrews <sup>1,2,4</sup> · Stanley Fields <sup>2,3</sup>

Received: 24 June 2020 / Revised: 18 January 2021 / Accepted: 25 January 2021 / Published online: 15 February 2021  
© The Author(s), under exclusive licence to International Society for Microbial Ecology 2021

## Abstract

As hosts acquire resistance to viruses, viruses must overcome that resistance to re-establish infectivity, or go extinct. Despite the significant hurdles associated with adapting to a resistant host, viruses are evolutionarily successful and maintain stable coevolutionary relationships with their hosts. To investigate the factors underlying how pathogens adapt to their hosts, we performed a deep mutational scan of the region of the  $\lambda$  tail fiber tip protein that mediates contact with the receptor on  $\lambda$ 's host, *Escherichia coli*. Phages harboring amino acid substitutions were subjected to selection for infectivity on wild type *E. coli*, revealing a highly restrictive fitness landscape, in which most substitutions completely abrogate function. A subset of positions that are tolerant of mutation in this assay, but diverse over evolutionary time, are associated with host range expansion. Imposing selection for phage infectivity on three  $\lambda$ -resistant hosts, each harboring a different missense mutation in the  $\lambda$  receptor, reveals hundreds of adaptive variants in  $\lambda$ . We distinguish  $\lambda$  variants that confer promiscuity, a general ability to overcome host resistance, from those that drive host-specific infectivity. Both processes may be important in driving adaptation to a novel host.

## Introduction

Viruses and their hosts engage in an evolutionary battle: mutations in the virus that increase infectivity come at a cost to bacterial growth, and mutations in the host that confer resistance to viruses come at a cost to the virus. In phages, as in other viruses, much of this battle is centered on the binding relationship between the host receptor and the viral protein that contacts it. Despite the marked adaptability of viruses, they face significant evolutionary hurdles. Their host populations rapidly evolve resistance, which may be

due to missense mutations in their receptors that disrupt binding or total loss of nonessential receptors. Thus, viruses must rapidly adapt to optimize weakened binding relationships and/or establish new relationships. In addition, viruses may face a polymorphic host population, which may lead to tradeoffs between viral generalism and specificity. Given these evolutionary hurdles, and given that bacteria frequently outcompete phages and drive them to extinction in co-culture, there remains the need to explain why viruses are so successful over evolutionary time scales [1–4].

Phages such as  $\lambda$  serve as a useful model for analyzing host-pathogen coevolution [3–5]. Co-culture of  $\lambda$  with its *Escherichia coli* host has resulted in the isolation of  $\lambda$ -resistant *E. coli* strains, which typically harbor null mutations of a maltose porin, LamB, that serves as the  $\lambda$  receptor [6–8]. Selection for both  $\lambda$ -resistance and maltose uptake has revealed LamB missense variants that disrupt  $\lambda$  binding [9].  $\lambda$  variants can then arise that re-establish infection on these previously resistant strains [9, 10]. When co-cultured with *E. coli* not expressing LamB,  $\lambda$  strains have been isolated that switch to use the noncanonical receptor OmpF, a LamB paralog [11].

Phage tail fibers, which bind to the host, evolve extremely quickly due to strong selective pressures and dedicated diversification mechanisms [12–14]. The  $\lambda$  tail fiber J protein consists of 1132 amino acids, with high conservation across its N-terminal ~980 amino acids and extreme

---

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41396-021-00912-2>.

✉ Stanley Fields  
fields@uw.edu

- <sup>1</sup> Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA
- <sup>2</sup> Department of Genome Sciences, University of Washington, Seattle, WA, USA
- <sup>3</sup> Department of Medicine, University of Washington, Seattle, WA, USA
- <sup>4</sup> Present address: Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, USA

diversification across its ~150 amino acid C-terminal domain, which contacts the receptor [8, 10]. As examples, the J proteins from two recently isolated phages, lambda\_2H10 and lambda\_2G7b, are each >97% identical to  $\lambda$  across their N-terminal 982 amino acids, but only 40–60% identical across their C-terminal ~150 residues [15]. The isolates are similarly diverged from each other and have unique host ranges. Although the structure of J or a homologous tail fiber protein has not been solved, tail fibers such as that of T4 form highly intertwined trimers [16, 17]. The trimer is composed almost entirely of  $\beta$ -sheets arranged in helical barrels, with the distal tip forming a more globular structure that directly contacts the receptor. Consistent with this structure, computational analysis of the secondary structure of J predicts  $\beta$ -sheets across the C-terminal domain, with a single  $\alpha$ -helix ~65 residues from the C-terminus. In T4, the tail fiber turns back on itself to form a six-sheeted rather than three-sheeted barrel, suggesting that the  $\alpha$ -helix may be part of the distal tip of J.

We decided to interrogate  $\lambda$  host range by generating a library of thousands of genetic variants in the C-terminal domain of J. We first imposed selection for infectivity on wild type (wt) *E. coli*, comparing the patterns of infectivity with the evolution and possible structure of the J protein. We then imposed selection on the same library for infectivity on three  $\lambda$ -resistant hosts, uncovering shared and unique paths to adaptation. By comparing variants within and between conditions, we consider the likely routes by which  $\lambda$  may overcome host resistance, and we ascribe distinct roles for variants that increase promiscuity from those that drive changes to specificity.

## Methods

### Generation of $\lambda$ variants by codon replacement

We generated a library consisting of most possible single amino acid substitutions in the C-terminal 150 codons of J. We first cloned the wild type J gene onto a high-copy plasmid and used site-directed mutagenesis to disrupt a single BbvCI site in the coding sequence, leaving two same-orientation BbvCI sites on the backbone. We used the BbvCI nickase-based mutagenesis strategy described by Wrenbeck et al. [18] to randomly replace codons in the targeted region with “NNN”, using 150 degenerate oligonucleotides that each bear homology to sequence flanking a single codon [18]. We then inserted a 15 base pair barcode downstream of J and generated a barcode–sequence map by paired-end Illumina sequencing of the mutagenized region, with the barcode on a separate indexing read. We ignored reads with no barcode inserted, barcodes with fewer than five high-quality reads, and barcodes linked to sequences

for which more than one read disagreed with the consensus at a given base pair. The barcoded and mapped sequence variants were then cut out of the plasmid backbone using *PasI* and ligated into *PasI*-digested and dephosphorylated  $\lambda$  DNA ( $\lambda_{\text{Sam7/cl857}}$ , NEB #N3011L, Ipswich, MA, USA). This strain background is obligately lytic at 37 °C and has an amber-suppressible lysis system that can infect and kill *E. coli*, but not release progeny, unless the strain is an amber suppressor [19]. The ligated DNA was packaged into virions using the MaxPlax  $\lambda$  packaging extract (Lucigen #MP5120, Middleton, WI, USA), and the virions were sequenced to determine input barcode frequency.

### Selection for infectivity

Prior to each selection, host cells were grown to  $\text{OD}_{600} = 0.3\text{--}0.4$  in LB supplemented with 10 mM  $\text{MgSO}_4$  and 0.2% maltose, and then pelleted. Approximately  $10^7$  virions were mixed with  $\sim 2 \times 10^8$  cells in the same medium in 1 mL total volume and incubated at 37 °C with shaking at 600 rpm. This binding step was 10 min for the wild type host (DH10B) [20] or 60 min for hosts with exogenously expressed receptors (DH10B-lamB<sup>Δ</sup> + p44K-lamB-var). Following binding, the cells were spun down (12,700 rpm  $\times$  30 s) and resuspended in LB, and this washing step was repeated twice. The cells were then incubated at 37 °C with shaking at 600 rpm for an additional 90 min to allow infected cells to produce  $\lambda$  progeny, which remained trapped in the cell because DH10B does not suppress the Sam7 allele in  $\lambda$ . Cells were spun down, resuspended in Tris-saline-EDTA buffer, and lysed with a combination of lysozyme and bead beating. The lysate was diluted into Tris-saline-magnesium buffer, cleared by centrifugation, and passed through a 0.2  $\mu\text{m}$  filter to remove un-lysed cells. The selection was repeated for four rounds, each representing one additional infectious cycle, on the wild type host, or three rounds on hosts with exogenously expressed receptors. At each timepoint, including  $t = 0$ , the population size was measured by plating on LE392MP cells, which suppress the Sam7 allele, and barcodes were amplified from the population by qPCR for sequencing. Each selection was performed in triplicate.

### Sequencing and scoring of variants

At each timepoint, we deeply sequenced barcodes amplified from the phage populations. We counted only barcodes that were in the barcode–sequence map and had an input read count of  $\geq 20$  reads. For each barcode, we estimated the number of progeny produced by the average virion containing that barcode in an infectious cycle (hereafter, progeny). We developed a novel scoring approach, called Model-Bounded Scoring, that directly estimates growth

rate,  $r_{var}$ , as opposed to commonly used methods that calculate a wild type-normalized enrichment score [21] (see also, Supplementary methods). By this approach, we measure the population size of the library at each timepoint, and we generate a null model of the expected sequencing counts for each barcode given its starting frequency and assuming no growth. We then estimate the subpopulation size using the sum of the measured and modeled counts and calculate a growth rate by regressing the log subpopulation size over time. We calculate progeny as the exponential of the growth rate,  $e^{r_{var}}$ , with one progeny meaning the parent survived but did not produce offspring. After four infectious cycles, variant scores were highly correlated between replicates ( $r^2 = 0.996$ , Supplementary Fig. S1). In this dataset, Model-Bounded Scoring better segregates synonymous and nonsense variants, provides better agreement between replicates, and provides better agreement between different barcodes linked to the same protein sequence than enrichment-based scoring (Supplementary Fig. S2).

### Selection on resistant hosts

A deep mutational scan of LamB, the  $\lambda$  receptor, showed that many point mutations specific for  $\lambda$  resistance are in or proximal to loop L6, the presumptive  $\lambda$  binding site [22]. We chose three receptor mutations, LamB-R219H, LamB-T264I, and LamB-G267D, that confer  $\lambda$  resistance but allow maltodextrin transport, based on the rationale that these mutations specifically disrupt the  $\lambda$  binding site rather than disrupting stability of the receptor. In a *lamB*-deletion background, we generated a *lac*-inducible expression vector for the wild type and the three *lamB* alleles, which were expressed using 0.1 mM IPTG. We imposed selection for infectivity on the  $\lambda$  library on each of the four *E. coli* strains, and we estimated progeny using Model-Bounded Scoring.

## Results

### Mutational scanning strategy yields infectivity measurements for phages with thousands of J variants

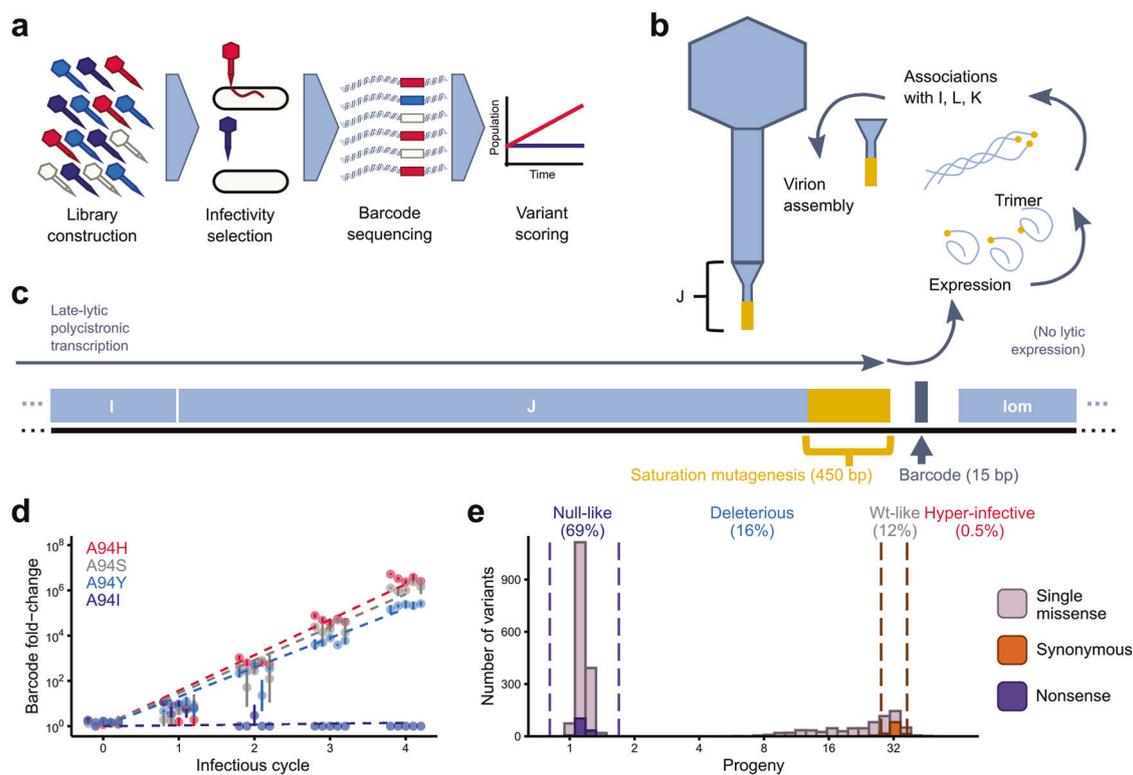
We performed a “deep mutational scan” of phage infectivity in  $\lambda$  by generating a library of phage variants with single amino acid changes in the tail fiber and imposing selection for infectivity on that library (Fig. 1a) [23]. The library consisted of nearly all single amino acid substitutions and several thousand double amino acid substitutions across the C-terminal 150 amino acids of J (Fig. 1b, c). For readability, we numbered the positions starting at the first residue of the mutagenized region. We imposed selection on this library by mixing the phage variants with wild type *E. coli* and

isolating intracellular phages after a single infectious cycle. This selection was repeated for four infectious cycles, and we estimated the abundance of each barcode in the expanding population over time using a novel approach we call “Model-Bounded Scoring” that directly estimates the progeny produced per generation. Barcode abundances between replicates showed strong agreement, as did different barcodes representing the same variant (Fig. 1d and Supplementary Fig. S2). We used the distribution of nonsense variants and synonymous variants to categorize variant effects. We considered variants with growth rate within two standard deviations of the mean nonsense variant to be “null-like” and those within two standard deviations of the mean synonymous variant to be “wt-like” (Fig. 1e). Variants falling between these two distributions were considered “deleterious”, while variants above the synonymous distribution were considered “hyper-infective”.

A large fraction (69%) of single missense variants conferred a null-like phenotype (Fig. 2a, Dataset 1). Although phage tails rapidly diversify to expand their host range, the fitness landscape of J with respect to its wild type host is much more restrictive than has been seen for other, much more conserved, proteins [24–27]. The pattern of mutational tolerance in J is broadly consistent with a T4-like barrel structure: extreme intolerance to substitutions to glycine and proline (<0.5% of substitutions to glycine and proline were wt-like and none was hyper-infective) and three regions of high periodicity in mutational tolerance, both of which suggest  $\beta$ -sheet richness (Fig. 2a and Supplementary Fig. S3). These patterns of mutational tolerance are not inconsistent with a coarse T4-like model of J’s structure (Fig. S4), with the receptor-binding portion of the protein encompassing positions ~50–100. This region of the protein was not, however, enriched for hyper-infective variants. We identified 12 such hyper-infective variants, two of which are accessible through point mutations. That these hyper-infective mutations are found throughout the mutagenized region suggests that residues that do not directly contact the receptor can nevertheless strongly influence receptor binding.

### J fitness landscape is more restrictive than predicted from evolutionary diversity

Because the rapid evolutionary diversification of phage tails appears to conflict with our observation of J’s broad intolerance to mutation, we decided to directly compare the patterns of mutational tolerance from our assay with patterns of mutational tolerance from an alignment of 910 orthologs of J (see Supplementary methods). However, we cannot exclude the possibility that our selection protocol was more stringent than natural selection, or that it imposed additional selective pressures by, for example, excluding



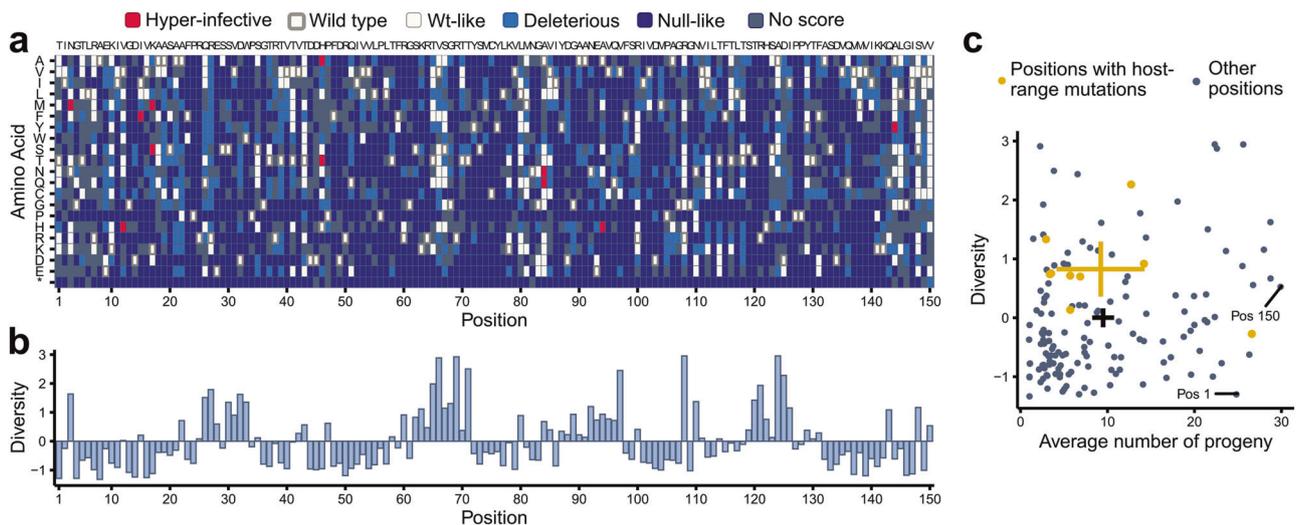
**Fig. 1 A sequencing-based method to assess phage infectivity across thousands of variants.** **a** The experimental strategy. A library of phage variants was constructed and subjected to selection for the ability to infect *E. coli*. After each infectious cycle, barcodes corresponding to each variant were deeply sequenced, and their frequencies were used to score each variant using Model-Bounded Scoring. **b** Binding of  $\lambda$  to the host is mediated by the J protein, which contacts the receptor, LamB. J spontaneously forms a homotrimer and then associates with accessory proteins that fold J into a mature conformation [44]. **c** J is the 3'-most gene on a long polycistronic transcript expressed in late lytic phase that contains most of the capsid genes. We mutagenized the 3'-most 450 bp of J (excluding the stop codon) using NNN codon replacement and inserted a 15 bp barcode downstream of the gene. **d** For each of four missense variants at A94, five randomly selected barcodes are plotted by their abundance in the phage population after each infectious cycle relative to the

pre-packaging DNA pool. Infectious cycle = 0 corresponds to the packaged but not yet selected phage population. Error bars represent the standard error between three replicate selections. For each variant, the growth rate,  $r$ , is the slope of an ordinary least squares regression line calculated separately for each barcode and averaged across all barcodes representing the same protein-level variant. The average growth rates for the selected variants are shown by slopes of the dashed lines. **e** Progeny per infectious cycle, equal to the exponential of the growth rate ( $e^r$ ), is shown for  $\lambda$  bearing synonymous (to wild type) variants in orange, nonsense variants in purple, and single missense variants in mauve. Progeny is always smaller than the theoretical burst size for a given variant because progeny is dependent on the probability of a virion binding during the binding step ( $\sim 0.5$  for wt  $\lambda$ ) and the efficiency of virion recovery following lysis. Dashed lines indicate score boundaries used to categorize variants as null-like, deleterious, wt-like, or hyper-infective.

weakly destabilized variants. Positions that harbored hyper-infective variants do not correspond to positions of high amino acid diversity across J orthologs (Fig. 2b). More broadly, the average progeny of variants at a position, a proxy of mutational tolerance, correlated only weakly ( $r^2 = 0.06$ ) with the evolutionary diversity of those positions (Fig. 2c, Dataset 2) [10, 11]. This observation contrasts with deep mutational scans of conserved cellular proteins, for which much stronger associations are observed between diversity and mutational tolerance [24, 25, 28]. One possible explanation for the discordance between fitness in our assay and fitness over evolutionary time is that our assay selected against destabilized variants because it contained a mechanical lysis step. However, variants were also observed that performed well in the assay despite not

having already fixed in  $\lambda$ ; for example, the best-scoring point mutation, I15F, was 4.2 standard deviations above the distribution of synonymous variants, producing 35% more progeny per generation than the wild type. In addition, the distribution of variant effects argues that J is not trapped on a fitness peak with respect to rapidly binding to its wild type host. Point mutations were slightly less likely to be strongly deleterious than all possible amino acid substitutions, even when synonymous mutations are excluded (Supplementary Fig. S5). Thus, the selection pressures that have shaped the evolutionary history and the wild type sequence of  $\lambda$  may have acted on a different property than is being selected for in our assays.

In J, peaks of diversity (Fig. 2b) at positions like 67 or 108, where many substitutions were wt-like, can be



**Fig. 2 Comparison of empirical mutational tolerance and evolutionary diversity.** **a** Categorical effects of all amino acid substitutions. Categories are defined in Fig. 1e. The wt sequence of the J region is shown across the top and the amino acid substitutions on the Y axis. **b** Amino acid diversity, calculated from ConSurf [45], across 910 orthologs of J. Zero represents the mean diversity across positions. **c** For each amino acid position, the evolutionary diversity (y-axis) is compared to the average progeny produced by amino acid substitutions at that position (x-axis). Diversity of each position correlates only

weakly with the average number of progeny ( $r^2 = 0.06$ ,  $p < 0.01$ ), in contrast to cellular proteins for which mutational tolerance and evolutionary diversity are more strongly related [24, 25, 28]. Gold points indicate positions where mutations have been reported that expand host range [10, 11]; these positions are more diverse but no more mutationally tolerant than the average position. Crosses represent 95% confidence intervals of the mean diversity and progeny of variants for positions with host range mutations (gold) or all positions (black). “Pos”: position.

distinguished from those at positions like 30 or 97, where most substitutions were deleterious or null-like. Furthermore,  $\lambda$  host range mutations tended to fall at these diverse, mutationally intolerant positions (gold points, Fig. 2c). We posit that by imposing selection for infectivity on a single host, we measured a selective pressure that is too narrow to adequately reflect  $\lambda$ 's entire evolutionary history on multiple hosts. Positions important for mediating infection of a single host would be mutationally intolerant in our assay, as the amino acid optimal for binding wild type LamB would likely have already been fixed in  $\lambda$ . However, these same positions could be under diversifying selection over evolutionary time during which the host has varied. Therefore, a complete understanding of  $\lambda$  evolution requires comparing the effects of J variants across many hosts.

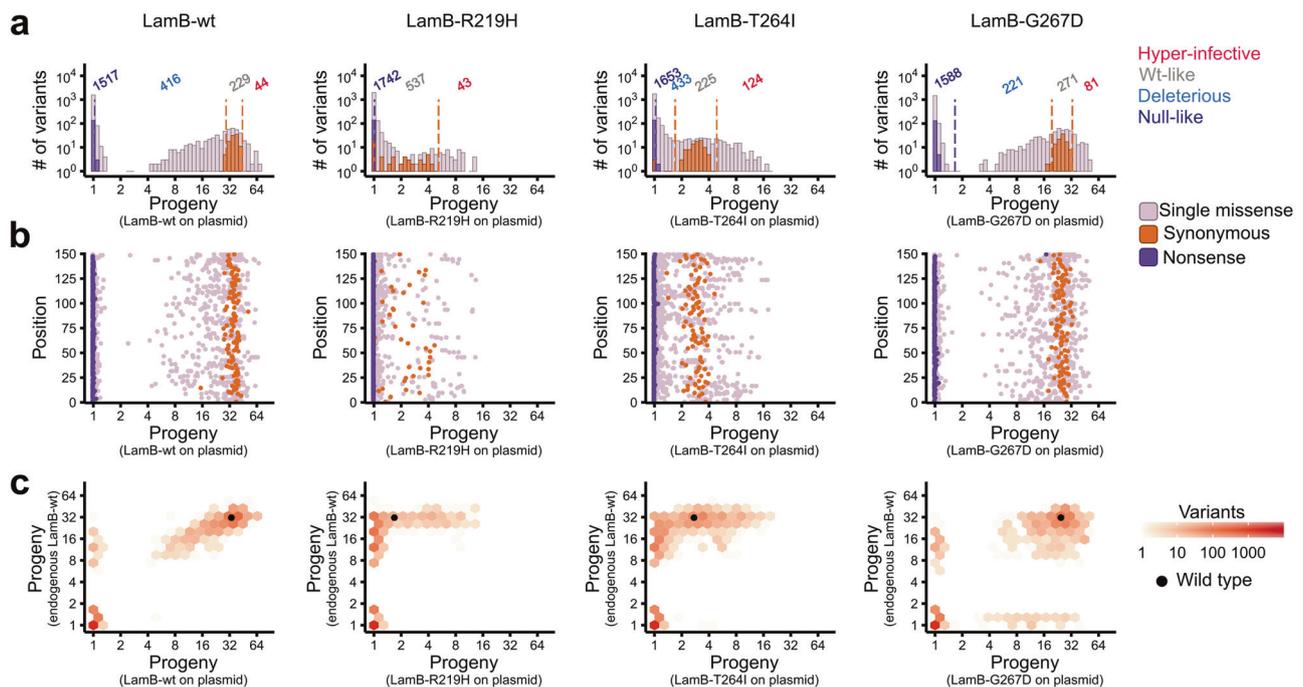
### Adaptation to a set of resistant hosts

To investigate mechanisms of adaptation to a resistant host, we challenged the library of J variants with a set of *E. coli* hosts bearing novel  $\lambda$  resistance mutations. Similar to selection on wild type *E. coli*, the average synonymous J variant was more infective than the average nonsense variant on each of the three resistant *E. coli* strains (Fig. 3a). Because Model-Bounded Scoring estimates real growth rate, rather than relative fitness, we can directly compare the growth of  $\lambda$  bearing the same J variants on different hosts without normalizing to a reference allele.  $\lambda$  was slightly

more infective on the IPTG-expressed wild type receptor than when this receptor was expressed from the endogenous *lamB* locus (31 progeny vs. 27 progeny for wild type J), but infectivity of the variants was highly correlated between the selections ( $r^2 = 0.958$ ). The resistant *lamB* alleles did not confer absolute resistance, but decreased the progeny produced by wild type  $\lambda$  from 31 (LamB-wt) to 22 (LamB-G267D), 2.7 (LamB-T264I), and 2.0 (LamB-R219H). For each *lamB* mutant, we could identify single missense variants in J that restored progeny to a substantial proportion of that produced on the wild type host. On LamB-R219H, the most resistant host, variants produced up to 40% of the progeny compared to that which wild type  $\lambda$  produced on LamB-wt (Fig. 3a). Moreover, a much larger fraction of variants significantly outperformed the synonymous distribution on each resistant host. Adaptive variants in each case were broadly distributed over the sequence, not highlighting a domain or structural feature uniquely necessary for adaptation (Fig. 3b). With some exceptions, adaptive variants were neutral-to-beneficial at infecting LamB-wt (Fig. 3c). However, the reverse was not true; variants that were highly infective on the wild type receptor were frequently less infective on the novel hosts.

### Specific and general mechanisms of adaptation

Comparing the infectivity of J variants between two hosts shows that the most infective J variant with respect to one



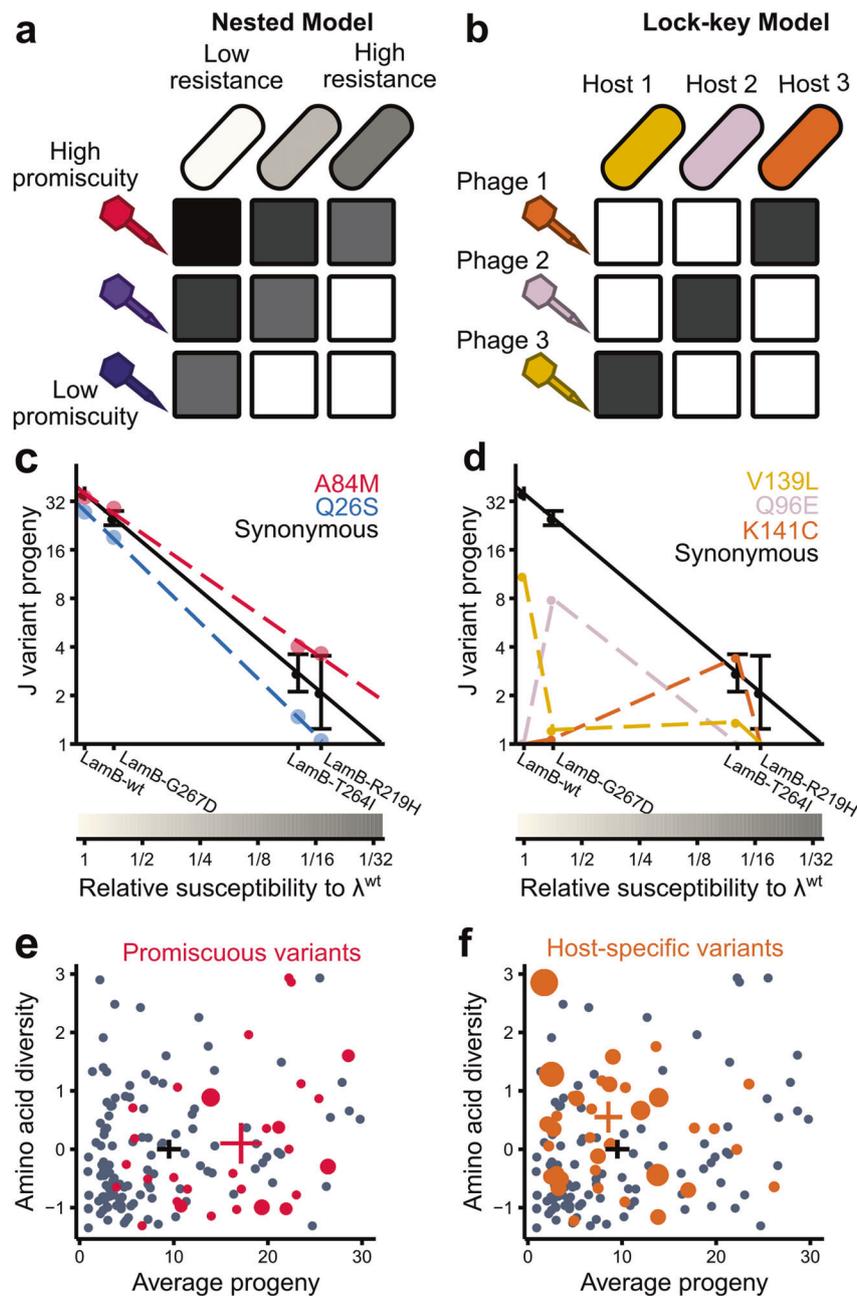
**Fig. 3 Selection of  $\lambda$  bearing J variants on  $\lambda$ -resistant hosts. a** Distribution of fitness effects on each host, expressed in terms of mean progeny produced per infectious cycle. In each case, the synonymous and nonsense distributions can be separated, despite the nominal  $\lambda$ -resistance of each lamB allele. Variants that outperform wild type  $\lambda$  do so by a larger margin on hosts that are more resistant. **b** Progeny for each J variant shown by the position of the mutation in the sequence. Adaptive mutations occur frequently at a subset of positions, but these positions are spread over the entire mutagenized region. Positions with

many hyper-infective mutations are more apparent on the more resistant hosts LamB-R219H and LamB-T264I than on LamB-wt or LamB-G267D. **c** Correlation between progeny produced by  $\lambda$  bearing each J variant on its wild type host (y-axis), or on a host bearing a plasmid-borne lamB allele (x-axis). Variants that are highly infective on a non-wt host tend to also be infective on the wild type host, with some exceptions. However, many variants that are highly infective on the wild type host are poorly infective on non-wt hosts.

host was frequently infective on the other host as well, with Spearman's  $\rho$  between variants' scores on different hosts ranging from  $\rho = 0.49$ – $0.56$ , depending on the hosts (Fig. 3c). In cases for which a J variant conferred infectivity on one host but not the other, it nearly always conferred infectivity on the host that is less resistant to wild type  $\lambda$  (Supplementary Figs. S6 and S7). Reciprocally, for hosts with greater resistance to wild type  $\lambda$ , a smaller fraction of J variants conferred infectivity (i.e., were not null-like). This pattern is consistent with a “nested” model in which each host and pathogen has a set amount of resistance or counter-resistance that is not dependent on the other player (Fig. 4a). In this model, infectivity is determined by the relative strength of resistance and counter-resistance. This model would contrast with a “lock–key” model [29], in which infectivity is determined by how well a given  $\lambda$  variant matches a given receptor (Fig. 4b). The relevant distinction is that under a nested model, but not a lock–key model, a J variant may gain generic counter-resistance that improves its infectivity on many hosts.

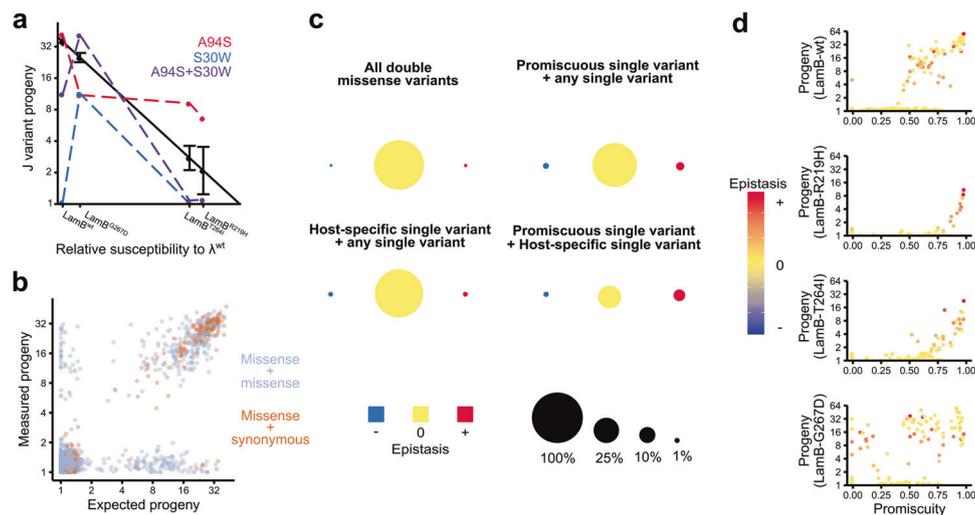
The property of generic counter-resistance can be compared to the property of enzyme “promiscuity”, in

which enzymes weakly catalyze noncanonical reactions in addition to their normal biological activities [30]. Under the right selection pressure, mutations can expand or improve these promiscuous activities, and evolution may eventually lead to specialization in the new activities [30]. By analogy, J might be said to have a baseline promiscuity in that it can weakly bind non-wt LamB receptors, and that variants have increased promiscuity if they improve infectivity broadly over the space of potential hosts relative to the most susceptible host, LamB-wt. For example, A84M (Fig. 4c, red line) has wt-like infectivity on the LamB-wt host, but it has a shallower slope across the set of resistant hosts compared to synonymous variants (Fig. 4c, black line). We calculated normalized Shannon's entropy, which measures the diversity of the hosts bound by a phage variant (see Supplementary methods), calling variants with greater entropy than any synonymous variant “promiscuous.” Although variants that are null-like on all hosts may have high entropy (Supplementary Fig. S8), we filtered out variants that were null-like on all hosts on the grounds that they could not be considered promiscuous. We observed a strong



**Fig. 4 Discrimination of promiscuity from specificity by comparison of variant effects across different hosts.** **a** Under a nested model, phages have more or less ability to generally infect hosts, which we describe as “promiscuity”, which contends with the level of resistance of potential hosts. **b** Under a lock-key model, the specific relationship between a host and phage determines infectivity, rather than host-independent properties of the phage or phage-independent properties of the host. **c** We can estimate the level of resistance of a given host by asking how well  $\lambda^{wt}$  produces progeny on it compared to LamB-wt ( $x$ -axis). Some J variants, like A84M are less affected by resistance than wild type  $\lambda$ , whereas others, like Q26S, are more affected. The black line represents synonymous variants, with error bars equal to  $\pm 1$  standard deviation. We calculate promiscuity as the normalized Shannon’s entropy for each variant across the four potential hosts. **d** Some J variants confer infectivity on only a single

host and are null-like on all others. Most of these variants, like Q96E, are deleterious, even on the host for which they are specific. Therefore, most of these variants cannot drive adaptation to a novel host by themselves, though they may work in concert with other variants. **e** Positions with promiscuous variants are shown in red with respect to their tolerance to mutation and amino acid diversity, with the size of the circle representing the number of unique promiscuous variants. The weighted average of these positions (cross, red) is more tolerant to mutation but not more diverse than the average of all positions (cross, black). Crosses represent 95% confidence intervals. **f** Variants that display specific infectivity on a single LamB variant fall at positions shown in orange. The weighted average of these positions has higher amino acid diversity, but is not more mutationally tolerant, than the average position.



**Fig. 5 Double missense variants in J can mediate adaptation to new hosts.** **a** The double missense variant S30W, A94S is a combination of a promiscuous variant (A94S) and a LamB-G267D-specific variant (S30W). The double missense variant exhibits sign epistasis, improving infectivity on LamB-G267D despite the deleteriousness of each single missense variant. **b** For double missense variants (silver), we calculated the expected progeny from the progeny of each of the single missense variants using a simple multiplicative model. A subset of double missense variants strongly deviates from the multiplicative model, in contrast to variants in which a single missense mutation is paired with a single synonymous mutation (orange), which are more likely to agree with the multiplicative model ( $r^2 = 0.92$  missense  $\times$

synonymous vs.  $r^2 = 0.64$  missense  $\times$  missense). **c** Across the four hosts, most double missense variants in J do not exhibit significant epistasis (top left panel). However, double missense variants that contain a promiscuous variant, a host-specific variant, or both, are more likely to exhibit significant epistasis. We measured significant epistasis in 13.2% of double missense variants containing promiscuous variants, 5.4% containing host-specific variants, and 23.8% containing both, compared to 1.8% containing neither. **d** For each host, progeny is positively associated with promiscuity and with positive epistasis. However, these associations become more salient on resistant hosts, with all infective variants on LamB-R219H being promiscuous and positively epistatic, compared to a minority of variants on LamB-wt.

positive relationship between promiscuity and growth on the wild type host (Supplementary Fig. S8), which implies that these promiscuous variants could persist in a  $\lambda$  population prior to encountering a resistant host. However, that these variants have not already been fixed in the population may imply subtle costs to promiscuity, such as thermodynamic instability [30, 31].

In addition, 49 single missense variants were capable of growth on only a single host, most of which (such as Q96E in Fig. 4d) conferred growth on LamB-G267D. With a few exceptions, like K141C (Fig. 4d), these host-specific variants were neither neutral nor beneficial on any host tested; they were merely dramatically more infective on one particular host compared to the other hosts. Thus, while these host-specific variants may be an important part of adapting to a host, they are generally insufficient to directly overcome resistance. Rather, we posit that the acquisition of host-specific variants may follow after the acquisition of promiscuous variants and either ameliorate thermodynamic costs or prevent off-target binding. This process could explain why very few LamB-wt-specific variants were observed, and why they were less infective than variants specific to other hosts (Supplementary Fig. S9), as if the wild type J sequence has already been selected for near-maximal specificity to its host.

Based on our analysis of J variants growing on a wild type host, positions with low mutational tolerance for growth on the wild type host but high diversity over evolutionary time are predicted to be enriched within variants that drive adaptation to a novel receptor (Fig. 2c). Promiscuous variants, however, tended to occur at positions with higher than average progeny across variants (Fig. 4e). By contrast, host-specific variants tended to occur at positions that are intolerant to mutation and diversified over evolutionary time (Fig. 4f). Thus, while our hypothesis that these mutationally intolerant, evolutionarily diverse positions are driving host-specificity is largely supported, host-specificity is not equivalent to overcoming resistance, which can happen through host-nonspecific mechanisms (i.e., promiscuity). This distinction implies that host range mutations arising in experimental evolution studies have not generally been promiscuous variants, as these host range mutations have mostly fallen at positions both mutationally intolerant and diverse (Fig. 2c). In our analysis, many of these host range mutations were null-like on all the hosts tested, including LamB-wt (Supplementary Table S1), suggesting that the effects of these variants may be specific to the hosts used in those studies and/or co-occurring mutations in J, or that our assay excluded weakly stable variants in a way that these other studies did not.

## Positive epistasis potentiates adaptation to a new host

In addition to single missense variants, the J library contained ~7500 variants with two missense mutations, sparsely surveying the space of >4 million possible double missense variants. We wondered whether combinations of promiscuous and host-specific mutations could help mediate adaptation to a novel receptor beyond what either class of mutations would confer in isolation. For example, S30W is a LamB-G267D-specific variant that was mildly impaired on LamB-G267D but null-like on the other receptors (Fig. 5a). Mutations at S30 have been associated with “the birth of clades” in J [32], and S30 is among the positions with the lowest tolerance to mutation relative to its evolutionary diversity (Supplementary Fig. S10). A94S, by contrast, is a mutation that has been observed in experimental evolution studies on host range expansion [10], but only in combination with other mutations. We observed that A94S conferred high promiscuity but was slightly deleterious on the LamB-G267D host. Despite the deleteriousness of S30W alone, the double missense variant A94S combined with S30W had both high growth and moderate specificity on LamB-G267D. This double missense variant therefore exhibits positive epistasis on one host, though it is poorly infective on the other hosts.

On LamB-wt, only 6.2% of double missense variants were infective (i.e., not null-like), compared to 26% of single missense variants that were infective (Dataset 3). For each double missense variant, we calculated the expected progeny given the progeny from each single variant, using a simple multiplicative model (see Supplementary methods). Most variants yielded similar values for measured and expected progeny, but ~45% of infective variants (2.7% of total variants) exhibited significant epistasis (Fig. 5b), including 22 infective variants for which both mutations were null-like on their own. These variants with dramatic reciprocal sign epistasis were enriched for positions with host range mutations [10, 11], which appeared 20 times among these 22 variants. As a control, we also considered variants in which a single missense mutation was paired with a synonymous mutation. Such variants exhibited better agreement between expected and empirical scores ( $r^2 = 0.92$ ) than double missense variants ( $r^2 = 0.64$ ), suggesting that the prevalence of epistasis in double missense variants is not an artifact of the fact that most double missense variants were less abundant in the library than single missense variants.

Similar to the selection on the wild type host, strong positive epistasis was prevalent on each resistant host. Single missense variants that conferred promiscuity frequently interacted epistatically with other variants (Fig. 5c). Of double missense variants that contained a promiscuous single missense variant, 90/676 (13.3%) exhibited significant

epistasis compared to only 2.2% of all double missense variants. The rare cases of double missense variants consisting of one promiscuous variant and one host-specific variant were even more likely to exhibit epistasis (6/25, 24%). Of these examples, 5/6 were in the positive direction.

Additionally, the effects of positive epistasis became more salient when  $\lambda$  was challenged with a resistant host. Double missense variants that were infective on LamB-wt had varying levels of promiscuity and positive epistasis (Fig. 5d). By contrast, double missense variants that were infective on the most resistant host, LamB-R219H, nearly exclusively had both high promiscuity and positive epistasis. The other two hosts revealed intermediate effects.

## Discussion

To investigate how  $\lambda$  overcomes host resistance, we analyzed thousands of variants of its tail fiber protein, J, on a small set of resistant *E. coli* hosts. We find that promiscuous J variants, which increase infectivity on a broad range of hosts, underlie the re-establishment of infection on resistant hosts. These variants co-exist with other, host-specific, variants that generally do not increase infectivity on any host but have smaller losses to infectivity on a single host. We posit that both types of variants are important for a phage to adapt to a new host, with host-specific variants likely ameliorating costs associated with promiscuity. This framing has implications for experimental evolution studies, protein adaptation more broadly, and natural phage-bacteria communities.

When phages and bacteria cyclically develop resistance and counter-resistance in experimental evolution studies, this coevolution is frequently characterized by an initial escalation of both host resistance and phage counter-resistance. This escalation eventually reaches an asymptote and is followed by negative frequency-dependent selection (“Kill the winner” dynamics) in which the dominant phages are most infective on the most common hosts [3, 33]. This pattern is well explained by a model in which promiscuous variants drive broadened host range but come at a cumulative cost, manifesting as lower growth rate relative to their host-specific counterparts [34]. The sequential acquisition of promiscuous variants could also open up pathways for a phage to infect a highly resistant host by first adapting to a less resistant host in the same environment. For example, Werts et al. [10] could not directly isolate  $\lambda$  that overcame the resistance allele LamB-G151D, but by pre-adapting the phage to other hosts with weaker resistance, they found double mutants able to grow on LamB-G151D [10]. However, thermodynamic costs associated with promiscuity may also constrain paths to adaptation. In their isolation of LamB-independent  $\lambda$  strains that use OmpF as receptor, Meyer et al. [11] repeatedly identified the same set of 4–5 adaptive variants in J across

dozens of independent cultures, suggesting considerable constraint on the path to counter-resistance [35]. In a follow-up study, the bi-specific J intermediate, which binds both LamB and OmpF, was less stable than the LamB-specific parent, and selection for OmpF specificity was sufficient to restore stability [36]. The adaptive variants identified by Meyer et al. [11] occur at positions that define clades of J [32], with some of these positions identified in our selection as being high in diversity but intolerant to mutation (Supplementary Fig. S10). These positions also correspond to some host-specific J variants, such as S30W (Fig. 5a).

Promiscuous activities provide convenient starting points for adaptation to novel protein functions [30]. The low level of infectivity mediated by wild type J on resistant hosts serves as a starting point for adaptation to those hosts by mutations that increase promiscuity. Second mutations can compensate for these initial mutations that confer promiscuity, offsetting their potential costs and resulting in highly infective and/or promiscuous double variants. This effect can be seen in the strong association between positive epistasis and promiscuity, whereby promiscuous single variants were more likely to have positive epistatic interactions than non-promiscuous variants (Fig. 5c). In addition, rare promiscuous variants that drive adaptation to LamB-R219H were always positively epistatic (Fig. 5d). At a mechanistic level, promiscuous J variants may shift between multiple semi-stable protein conformations, or they may heterogeneously fold into one of multiple stable conformations. The latter mechanism was found to underlie the LamB-OmpF bi-specific intermediates of J characterized by Petrie et al. [36]. Under a model requiring multiple protein conformations, destabilization may be fundamentally linked to promiscuity rather than incidental to it. Therefore, counter to observations with promiscuous enzymes for which stabilization precedes and potentiates adaptive variation [31, 37], stabilizing mutations in phages are unlikely to precede adaptive ones. Instead, a destabilizing mutation that increases promiscuity must come first, followed by a compensatory mutation that re-stabilizes the protein into an optimal conformation for infection of the most abundant host. This prediction would also explain why J is so broadly intolerant of mutation: repeated evolutionary transitions between stable and unstable sequences leave J close to a threshold of severe destabilization [38, 39].

Naturally occurring phage–bacteria interactions also show patterns consistent with a balance between host-specific and promiscuous variation. Phages and bacteria isolated from the same environment form infection networks exhibiting both nestedness and “modularity”, a property of lock–key models, with nestedness dominating at small scales involving highly related strains [40, 41]. This pattern is consistent with promiscuity driving counter-resistance to a newly resistant host, and modularity arising

between more diverged hosts. The extent to which the evolution of phages involves adaptation to diverged hosts remains unclear. Orthologs of both J and LamB are broadly distributed among enterobacteria, and even appear in distant  $\epsilon$ -proteobacteria species, suggesting either an ancient origin of this host–pathogen relationship or frequent cross-taxa jumps in host range. However, the limited breadth of hosts in which nestedness is observed suggests practical constraints to the promiscuity of a phage tail: a single promiscuous phage variant is more likely to be infective on multiple receptor variants within a single host species than on receptors of multiple related host species.

We conclude that although  $\lambda$  faces significant evolutionary hurdles not faced by its host, it can establish common paths to adaptation on multiple potential hosts by maintaining a balance between promiscuity and host-specificity. This balance may be mediated by mild destabilization of the protein, allowing it to sample multiple conformations, although further work is needed to directly test this hypothesis. Although we surveyed adaptation to only a small set of resistant hosts, this general framework is consistent with prior observations of how  $\lambda$  evolves to switch to a novel receptor. This framework may apply broadly to other phages and viruses for which mutations are more difficult to assay en masse [33, 42, 43].

**Acknowledgements** We thank Benjamin Kerr for his advice on the manuscript, and Alan Rubin and Matt Rich for scripts that facilitated filtering and counting barcode sequences. This work was supported by grants T32 HG000035 and RM1 HG010461 from the National Human Genome Research Institute of the NIH.

## Compliance with ethical standards

**Conflict of interest** The authors declare no competing interests.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

1. Lenski RE. Coevolution of bacteria and phage: are there endless cycles of bacterial defenses and phage counterdefenses? *J Theor Biol.* 1984;108:319–25.
2. Lenski RE, Levin BR. Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. *Am Nat.* 1985;125:585–602.
3. Koskella B, Brockhurst MA. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev.* 2014;38:916–31.
4. Dennehy JJ. What can phages tell us about host-pathogen coevolution? *Int J Evol Biol.* 2012;2012:396165.
5. Buckling A, Rainey PB. Antagonistic coevolution between a bacterium and a bacteriophage. *Proc R Soc B Biol Sci.* 2002;269:931–6.
6. Lederberg E. Pleiotropy for maltose fermentation and phage resistance in *Escherichia coli* K-12. *Genetics.* 1955; 40:580–1.

7. Thirion JP, Hofnung M. On some genetic aspects of phage  $\lambda$  resistance. *Genetics*. 1972;71:207–16.
8. Gehring K, Charbit A, Brissaud E, Hofnung M. Bacteriophage lambda receptor site on the *Escherichia coli* K-12 LamB protein. *J Bacteriol*. 1987;169:2103–6.
9. Shaw JE, Bingham H, Fuerst CR, Pearson ML. The multisite character of host-range mutations in bacteriophage  $\lambda$ . *Virology*. 1977;83:180–94.
10. Werts C, Michel V, Hofnung M, Charbit A. Adsorption of bacteriophage lambda on the LamB protein of *Escherichia coli* K-12: point mutations in gene J of lambda responsible for extended host range. *J Bacteriol*. 1994;176:941–7.
11. Meyer JR, Dobias DT, Weitz JS, Barrick JE, Quick RT, Lenski RE. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science*. 2012;335:428–32.
12. Doulatov S, Hodes A, Dal L, Mandhana N, Liu M, Deora R, et al. Tropism switching in *Bordetella* bacteriophage defines a family of diversity-generating retroelements. *Nature*. 2004;431:476–81.
13. Tetart F, Desplats C, Krisch HM. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. *J Mol Biol*. 1998;282:543–56.
14. Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, et al. Antagonistic coevolution accelerates molecular evolution. *Nature*. 2010;464:275–8.
15. Mathieu A, Dion M, Deng L, Tremblay D, Moncaut E, Shah SA, et al. Virulent coliphages in 1-year-old children fecal samples are fewer, but more infectious than temperate coliphages. *Nat Commun*. 2020;11:378.
16. Bartual SG, Otero JM, Garcia-Doval C, Llamas-Saiz AL, Kahn R, Fox GC, et al. Structure of the bacteriophage T4 long tail fiber receptor-binding tip. *PNAS*. 2010;107:20287–92.
17. Hyman P, van Raaij M. Bacteriophage T4 long tail fiber domains. *Biophys Rev* 2018;10:463–71.
18. Wrenbeck EE, Klesmith JR, Stapleton JA, Adeniran A, Tyo KEJ, Whitehead TA. Plasmid-based one-pot saturation mutagenesis. *Nat Methods*. 2016;13:928–30.
19. Kaiser D, Masuda T. In vitro assembly of bacteriophage Lambda heads. *PNAS*. 1973;70:260–4.
20. Durfee T, Nelson R, Baldwin S, Plunkett G, Burland V, Mau B, et al. The complete genome sequence of *Escherichia coli* DH10B: Insights into the biology of a laboratory workhorse. *J Bacteriol*. 2008;190:2597–606.
21. Rubin AF, Gelman H, Lucas N, Bajjalieh SM, Papenfuss AT, Speed TP, et al. A statistical framework for analyzing deep mutational scanning data. *Genome Biol*. 2017;18:150.
22. Andrews B, Fields S. Distinct patterns of mutational sensitivity for  $\lambda$  resistance and maltodextrin transport in *Escherichia coli* LamB. *Microb Genom*. 2020;6:e000364.
23. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. *Nat Methods* 2014;11:801–7.
24. Roscoe BP, Bolon DNA. Systematic exploration of ubiquitin sequence, E1 activation efficiency, and experimental fitness in yeast. *J Mol Biol*. 2014;426:2854–70.
25. Mishra P, Flynn J, Starr TN, Bolon DNA. Systematic mutant analyses elucidate general and client-specific aspects of Hsp90 function. *Cell Rep*. 2016;15:588–98.
26. Bandaru P, Shah NH, Bhattacharyya M, Barton JP, Kondo Y, Cofsky JC, et al. Deconstruction of the ras switching cycle through saturation mutagenesis. *Elife*. 2017;6:e27810.
27. Weile J, Sun S, Cote AG, Knapp J, Verby M, Mellor JC, et al. A framework for exhaustively mapping functional missense variants. *Mol Syst Biol*. 2017;13:957.
28. Boucher JI, Cote P, Flynn J, Jiang L, Laban A, Mishra P, et al. Viewing protein fitness landscapes through a next-gen lens. *Genetics*. 2014;198:461–71.
29. Weitz JS, Hartman H, Levin SA. Coevolutionary arms races between bacteria and bacteriophage. *PNAS*. 2005;102:9535–40.
30. Arnold FH. How proteins adapt: lessons from directed evolution. *Cold Spring Harb Symp Quant Biol*. 2009;74:41–6.
31. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. *PNAS*. 2006;103:5869–74.
32. Maddamsetti R, Johnson DT, Spielman SJ, Petrie KL, Marks DS, Meyer JR. Gain-of-function experiments with bacteriophage lambda uncover residues under diversifying selection in nature. *Evolution*. 2018;72:2234–43.
33. Hall JPJ, Harrison E, Brockhurst MA. Viral host-adaptation: insights from evolution experiments with phages. *Curr Opin Virol*. 2013;3:572–7.
34. Poullain V, Gandon S, Brockhurst MA, Buckling A, Hochberg ME. The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage. *Evolution*. 2007;62:1–11.
35. Franke J, Kloser A, De Visser JAGM, Krug J. Evolutionary accessibility of mutational pathways. *PLoS Comput Biol*. 2011;7:e1002134.
36. Petrie KL, Palmer ND, Johnson DT, Medina SJ, Yan SJ, Li V, et al. Destabilizing mutations encode nongenetic variation that drives evolutionary innovation. *Science*. 2018;359:1542–5.
37. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol*. 2009;19:596–604.
38. Gong LI, Suchard MA, Bloom JD. Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife*. 2013;2:e00631.
39. Tokuriki N, Tawfik DS. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature*. 2009;459:668–73.
40. Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS. Statistical structure of host-phage interactions. *PNAS*. 2011;108:E288–97.
41. Weitz JS, Poisot T, Meyer JR, Flores CO, Valverde S, Sullivan MB, et al. Phage-bacteria infection networks. *Trends Microbiol*. 2013;21:82–91.
42. Carrat F, Flahault A. Influenza vaccine: the challenge of antigenic drift. *Vaccine*. 2007;25:6852–62.
43. Liao HX, Lynch R, Zhou T, Gao F, Munir Alam S, Boyd SD, et al. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature*. 2013;496:469–76.
44. Rajagopala SV, Casjens S, Uetz P. The protein interaction map of bacteriophage lambda. *BMC Microbiol*. 2011;11:213.
45. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016;44:W344–50.