

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

Bryan Andrews<sup>1,2</sup> and Stanley Fields<sup>2,3,\*</sup>

1. Molecular and Cellular Biology Program, University of Washington, Seattle WA, USA
2. Department of Genome Sciences, University of Washington, Seattle WA, USA
3. Department of Medicine, University of Washington, Seattle WA, USA

\*corresponding author: Stanley Fields: [fields@uw.edu](mailto:fields@uw.edu), ph 206 616-4522

The authors declare that they have no conflicts of interest.

## 1 **Abstract**

2           **As hosts acquire resistance to viruses, viruses must overcome that resistance to re-establish infectivity,**  
3 **or go extinct. Despite the significant hurdles associated with adapting to a resistant host, viruses are**  
4 **evolutionarily successful and maintain stable coevolutionary relationships with their hosts. To investigate the**  
5 **factors underlying how pathogens adapt to their hosts, we performed a deep mutational scan of the region of**  
6 **the  $\lambda$  tail fiber tip protein that mediates contact with the  $\lambda$  host, *E. coli*. Phages harboring amino acid**  
7 **substitutions were subjected to selection for infectivity on wild type *E. coli*, revealing a highly restrictive fitness**  
8 **landscape, in which most substitutions completely abrogate function. By comparing this lack of mutational**  
9 **tolerance to evolutionary diversity, we highlight a set of mutationally intolerant and diverse positions**  
10 **associated with host range expansion. Imposing selection for infectivity on three  $\lambda$ -resistant hosts, each**  
11 **harboring a different missense mutation in the  $\lambda$  receptor, reveals hundreds of adaptive variants in  $\lambda$ . We**  
12 **distinguish  $\lambda$  variants that confer promiscuity, a general ability to overcome host resistance, from those that**  
13 **drive host-specific infectivity. Both processes may be important in driving adaptation to a novel host.**

14

## 15 **Introduction**

16           Viruses and their hosts engage in an evolutionary battle: mutations in the virus that increase infectivity  
17 come at a cost to bacterial growth, and mutations in the host that confer resistance to viruses come at a cost to  
18 the virus. In phages, as in other viruses, much of this battle is centered on the binding relationship between the  
19 host receptor and the viral protein that contacts it. Despite the adaptability of viruses to resistant hosts, viruses  
20 face significant evolutionary hurdles that their hosts do not (Fig. 1). First, random mutations are much more likely  
21 to disrupt an existing binding relationship than to generate a new one. Thus, viruses must survey a much broader  
22 sequence space than their hosts do to compete in an evolutionary arms race. Second, many receptors are readily  
23 lost in the host, but are essential to the virus. Some viruses, including  $\lambda$ , switch receptors following loss of their  
24 canonical receptor, although it is unclear if this is a common response [1, 2, 3]. Third, host populations are  
25 polymorphic, such that selection for viral resistance may give rise to multiple sub-clades with distinct resistance  
26 mechanisms. Overcoming any single resistance mechanism may not be sufficient for a virus to reestablish  
27 infectivity at the population level. Given these evolutionary hurdles, there remains the need to explain why viruses  
28 are so successful over evolutionary time scales [4, 5, 6]. In humans and other long-lived organisms, differences in  
29 relative generation times and mutation rates likely play a role [7]. However, phages are also extremely successful  
30 predators despite the short generation times and rapid adaptation of their bacterial hosts, which frequently  
31 outcompete phages and drive them to extinction in co-culture [6, 8].

32 Phages such as  $\lambda$  serve as a useful model for analyzing host-pathogen coevolution [6, 8, 9]. Co-culture of  
33  $\lambda$  with its *E. coli* host has resulted in the isolation of  $\lambda$ -resistant *E. coli* strains, which typically harbor null mutations  
34 of a maltose porin, LamB, that serves as the  $\lambda$  receptor [10, 11, 12]. Selection for both  $\lambda$ -resistance and maltose  
35 uptake has revealed LamB missense variants that disrupt  $\lambda$  binding [13].  $\lambda$  variants can then arise that can re-  
36 establish infection on these previously resistant strains [13, 14]. When co-cultured with *E. coli* not expressing  
37 LamB,  $\lambda$  strains have been isolated that switch to use the non-canonical receptor OmpF, a LamB paralog [3].

38 Phage tail fibers, which bind to the host, evolve extremely quickly due to strong selective pressures and  
39 dedicated diversification mechanisms [2, 15, 16]. The  $\lambda$  tail fiber J protein consists of 1,132 amino acids, with high  
40 conservation across its N-terminal ~980 amino acids and extreme diversification across its ~150 amino acid C-  
41 terminal domain, which contacts the receptor [12, 14]. As examples, the J proteins from two recently isolated  
42 phages, lambda\_2H10 and lambda\_2G7b, are each >97% identical to  $\lambda$  across the N-terminal 982 amino acids, but  
43 only 40-60% identical across the C-terminal ~150 residues [18]. The isolates are similarly diverged from each other  
44 and have unique host ranges. Although the structure of J or homologous tail fiber proteins has not been solved,  
45 tail fibers such as that of T4 form highly intertwined trimers [19, 20]. The trimer is composed almost entirely of  $\beta$ -  
46 sheets arranged in helical barrels, with the distal tip forming a more globular structure that directly contacts the  
47 receptor. Consistent with this structure, computational analysis of the secondary structure of J predicts  $\beta$ -sheets  
48 across the C-terminal domain, with a single  $\alpha$ -helix ~65 residues from the C-terminus. In T4, the tail fiber turns  
49 back on itself to form a six-sheeted, rather than three-sheeted, barrel, suggesting that the  $\alpha$ -helix may be part of  
50 the distal tip of J.

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

57

## 58 **Methods**

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

60 We generated a library consisting of most possible single amino acid substitutions in the C-terminal 150  
61 codons of J. We first cloned the wild type J gene onto a high-copy plasmid and used site-directed mutagenesis to  
62 disrupt a single BbvCI site in the coding sequence, leaving two same-orientation BbvCI sites on the backbone. We  
63 used the BbvCI nickase-based mutagenesis strategy described by Wrenbeck et al. (2016) to randomly replace

64 codons in the targeted region with ‘NNN,’ using 150 degenerate oligonucleotides that each bear homology  
65 flanking a single codon [21]. We then inserted a 15 base pair barcode downstream of J and generated a barcode-  
66 sequence map by paired-end Illumina sequencing of the mutagenized region, with the barcode on a separate  
67 indexing read. We ignored reads with no barcode inserted, barcodes with fewer than five high-quality reads, and  
68 barcodes linked to sequences for which more than one read disagreed with the consensus at a given base pair.  
69 The barcoded and mapped sequence variants were then cut out of the plasmid backbone using PstI and ligated  
70 into PstI-digested and dephosphorylated  $\lambda$  DNA ( $\lambda_{\text{Sam7/cl857}}$ , NEB #N3011L, Ipswich, MA, USA). This strain  
71 background is obligately lytic at 37°C and has an amber-suppressible lysis system that can infect and kill *E. coli*,  
72 but not release progeny, unless the strain is an amber suppressor [22]. The ligated DNA was packaged into virions  
73 using the MaxPlax  $\lambda$  packaging extract (Lucigen #MP5120, Middleton, WI, USA), and the virions were sequenced  
74 to determine input barcode frequency.

75

## 76 **Selection for infectivity**

77 For each selection,  $\sim 10^7$  virions we mixed with  $\sim 2 \times 10^8$  cells in rich media containing 10 mM  $\text{MgSO}_4$  in 1  
78 mL total volume and incubated at 37°C with shaking at 600 rpm. This binding step was 10 minutes for the wild  
79 type host (DH10B) [23] or 60 minutes for hosts with exogenously expressed receptors (DH10B-lamB<sup>A</sup> + p44K-lamB-  
80 var). Following binding, the cells were spun down (12,700 rpm x 30 s) and resuspended in LB, and this washing  
81 step was repeated twice. The cells were then incubated at 37°C with shaking at 600rpm for an additional 90  
82 minutes to allow infected cells to produce  $\lambda$  progeny, which remained trapped in the cell because DH10B does  
83 suppress the Sam7 allele in  $\lambda$ . Cells were spun down, resuspended in Tris-saline-EDTA buffer, and lysed with a  
84 combination of lysozyme and bead beating. The lysate was diluted into Tris-saline-magnesium buffer, cleared by  
85 centrifugation, and passed through a 0.2 mm filter to remove un-lysed cells. This procedure was repeated for four  
86 rounds on the wild type host, or three rounds on hosts with exogenously expressed receptors. At each timepoint,  
87 including t=0, the population size was measured by plating on LE392MP cells, which suppress the Sam7 allele, and  
88 barcodes were amplified from the population by qPCR for sequencing. Each selection was performed in triplicate.

89

## 90 **Sequencing and scoring of variants**

91 At each timepoint, we deeply sequenced barcodes amplified from the phage populations. We counted  
92 only barcodes that were in the barcode–sequence map and had an input read count of  $\geq 20$  reads. For each  
93 barcode, we estimated the number of progeny produced by the average virion containing that barcode in an  
94 infectious cycle (hereafter, progeny). We developed a novel scoring approach, called Model-Bounded Scoring,  
95 that directly estimates growth rate,  $r_{\text{var}}$ , as opposed to commonly used methods that calculate a wild type-

96 normalized enrichment score [24]. By this approach, we measure the population size of the library at each  
97 timepoint, and we generate a null model of the expected sequencing counts for each barcode given its starting  
98 frequency and assuming no growth. We then estimate the subpopulation size using the sum of the measured and  
99 modelled counts and calculate a growth rate by regressing the log subpopulation size over time. We calculate  
100 progeny as the exponential of the growth rate,  $e^{r_{var}}$ , with one progeny meaning the parent survived but did not  
101 produce offspring. After four infectious cycles, variant scores were highly correlated between replicates ( $r^2 =$   
102 0.996, Fig. S1). In this dataset, Model-Bounded Scoring better segregates synonymous and nonsense variants,  
103 provides better agreement between replicates, and provides better agreement between different barcodes linked  
104 to the same protein sequence than enrichment-based scoring (Fig. S2).

105

## 106 **Results**

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

108 We performed a ‘deep mutational scan’ of phage infectivity in  $\lambda$  by generating a library of phage variants  
109 with single amino acid changes in the tail fiber and imposing selection for infectivity on that library (Fig. 2a) [25].  
110 The library consisted of nearly all single amino acid substitutions and several thousand double amino acid  
111 substitutions across the C-terminal 150 amino acids of J (Fig. 2b, c). For readability, we numbered the positions  
112 starting at the first residue of the mutagenized region. We imposed selection on this library by mixing the phage  
113 variants with wild type *E. coli* and isolating intracellular phages after a single round of growth. This selection was  
114 repeated for four infectious cycles, and we estimated the abundance of each barcode in the expanding population  
115 over time using a novel approach we call ‘Model-Bounded Scoring’ that directly estimates the progeny produced  
116 per generation. Barcode abundances between replicates showed strong agreement, as did different barcodes  
117 representing the same variant (Fig. 2d, Fig. S2). We used the distribution of nonsense variants and synonymous  
118 variants to categorize variant effects. We considered variants with growth rate within two standard deviations of  
119 the mean nonsense variant to be ‘null-like’ and those within two standard deviations of the mean synonymous  
120 variant to be ‘wt-like’ (Fig. 2e). Variants falling between these two distributions were considered ‘deleterious,’  
121 while variants above the synonymous distribution were considered ‘hyper-infective.’

122 A large fraction (69%) of single missense variants conferred a null-like phenotype (Fig. 3a). Although phage  
123 tails rapidly diversify to expand their host range, the fitness landscape of J with respect to its wild type host is  
124 much more restrictive than has been seen for other, much more conserved, proteins [26, 27, 28, 29]. The pattern  
125 of mutational tolerance in J is broadly consistent with a T4-like barrel structure: extreme intolerance to  
126 substitutions to glycine and proline and three regions of high periodicity in mutational tolerance, both of which  
127 suggest  $\beta$ -sheet richness (Fig. 3a, S3). We can use this pattern of mutational tolerance to construct a coarse model

128 of J's structure (Fig. S4), with the receptor-binding portion of the protein encompassing positions ~50-100. This  
129 region of the protein is not, however, enriched for hyper-infective variants. We identified 12 such hyper-infective  
130 variants, two of which are accessible through point mutations. That these hyper-infective mutations are found  
131 throughout the mutagenized region suggests that residues that do not directly contact the receptor can  
132 nevertheless strongly influence receptor binding.

133

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

135 The extreme intolerance to mutation of J relative to other proteins conflicts with the observation that  
136 phage tail fibers, including J homologs, rapidly diversify over evolutionary time. We therefore decided to directly  
137 compare the patterns of mutational tolerance from our assay against patterns of mutational tolerance from an  
138 alignment of 910 orthologs of J. Positions that harbored hyper-infective variants do not correspond to positions  
139 of high amino acid diversity across J orthologs (Fig. 3b). More broadly, the average progeny of variants at a  
140 position, a proxy of mutational tolerance, correlated only weakly ( $r^2 = 0.06$ ) with the evolutionary diversity of  
141 those positions (Fig. 3c) [3, 14]. This observation contrasts with deep mutational scans of conserved cellular  
142 proteins, for which much stronger associations are observed between diversity and mutational tolerance [26, 27,  
143 30]. Additionally, the distribution of variant effects argues that J is not trapped on a fitness peak with respect to  
144 rapidly binding to its wild type host. Point mutations were slightly less likely to be strongly deleterious than all  
145 possible amino acid substitutions, even when synonymous mutations are excluded (Fig. S5). Moreover, the best-  
146 scoring point mutation, I15F, was 4.2 standard deviations above the distribution of synonymous variants,  
147 producing 35% more progeny per generation than the wild type. Thus, the selection pressures that have shaped  
148 the evolutionary history and the wild type sequence of  $\lambda$  may have acted on a different property than is being  
149 selected for in our assays.

150 In J, peaks of diversity (Fig. 3b) that corresponded to many wild type-like variants (*e.g.* positions 67 or 108)  
151 can be separated from those that corresponded to many deleterious or null-like variants (*e.g.* positions 30 and  
152 97). Furthermore,  $\lambda$  host range mutations tended to fall at these diverse, mutationally intolerant positions (gold  
153 points, Fig. 3c). We posit that by imposing selection for infectivity on a single host, we measured a selective  
154 pressure that is too narrow to adequately reflect  $\lambda$ 's entire evolutionary history on multiple hosts. Positions  
155 important for mediating infection of a single host would be mutationally intolerant in our assay, as the amino acid  
156 optimal for binding wild type LamB would likely have already been fixed in  $\lambda$ . However, these same positions could  
157 be under diversifying selection over evolutionary time during which the host has varied. Therefore, a complete  
158 understanding of  $\lambda$  evolution requires comparing the effects of J variants across many hosts.

159

## 160 **Adaptation to a set of resistant hosts**

161 To investigate mechanisms of adaptation to a resistant host, we challenged the library of J variants with  
162 a set of *E. coli* hosts bearing novel  $\lambda$  resistance mutations. A deep mutational scan of LamB, the  $\lambda$  receptor, showed  
163 that many point mutations specific for  $\lambda$  resistance are in or proximal to loop L6, the presumptive  $\lambda$  binding site  
164 [31]. We chose three receptor mutations, LamB-R219H, LamB-T264I, and LamB-G267D, that confer  $\lambda$  resistance  
165 but allow maltodextrin transport, on the basis that these mutations specifically disrupt the  $\lambda$  binding site rather  
166 than disrupting stability of the receptor. In a *lamB*-deletion background, we generated a *lac*-inducible expression  
167 vector for the wild type and the three *lamB* alleles, which were expressed using 0.1 mM IPTG. We imposed  
168 selection for infectivity on the  $\lambda$  library on each of the four *E. coli* strains, and we estimated progeny using Model-  
169 Bounded Scoring.

170 Similar to selection on wild type *E. coli*, synonymous J variants were more infective than nonsense variants  
171 on each of the three resistant *E. coli* strains (Fig. 4a). Because Model-Bounded Scoring estimates real growth rate,  
172 rather than relative fitness, we can directly compare the growth of  $\lambda$  bearing the same J variants on different hosts  
173 without normalizing to a reference allele.  $\lambda$  was slightly more infective on the IPTG-expressed wild type receptor  
174 than when this receptor was expressed from the endogenous *lamB* locus (31 progeny vs. 27 progeny for wild type  
175 J), but infectivity of the variants was highly correlated between the selections ( $r^2 = 0.958$ , Fig. 4c). The resistant  
176 *lamB* alleles did not confer absolute resistance, but decreased the progeny produced by wild type  $\lambda$  from 31  
177 (LamB-wt) to 22 (LamB-G267D), 2.7 (LamB-T264I), and 2.0 (LamB-R219H). For all three *lamB* mutants, we could  
178 identify single missense variants in J that restored progeny to 40-100% of what was produced on the wild type  
179 host (Fig. 4a). Moreover, a much larger fraction of variants significantly outperformed the synonymous  
180 distribution on each resistant host. Adaptive variants in each case were broadly distributed over the sequence,  
181 not highlighting a domain or structural feature uniquely necessary for adaptation (Fig. 4b). With some exceptions,  
182 adaptive variants were neutral-to-beneficial at infecting LamB-wt (Fig. 4c). However, the reverse was not true;  
183 variants that were highly infective on the wild type receptor were frequently less infective on the novel hosts.

184

## 185 **Specific and general mechanisms of adaptation**

186 A comparison of the infectivity of J variants between two hosts shows that the most infective J variant  
187 with respect to one host was frequently highly infective on the other host as well (Fig. 4c, S6). In cases for which  
188 a J variant conferred infectivity on one host but not the other, it nearly always conferred infectivity on the host  
189 that is less resistant to wild type  $\lambda$ . Reciprocally, for hosts with greater resistance to wild type  $\lambda$ , a smaller fraction  
190 of J variants conferred infectivity (*i.e.*, were not null-like). This pattern is consistent with a 'nested' model in which  
191 each host and pathogen has a set amount of resistance or counter-resistance that is not dependent on the other

192 player (Fig. 5a). In this model, infectivity is determined by the relative strength of resistance and counter-  
193 resistance. This model would contrast with a ‘lock-key’ model [34], in which infectivity is determined by how a  
194 well a given  $\lambda$  variant matches a given receptor (Fig. 5b). The relevant distinction is that under a nested model,  
195 but not a lock-key model, a J variant may gain generic counter-resistance that improves its infectivity on many  
196 hosts.

197 The property of generic counter-resistance can be compared to the property of enzyme ‘promiscuity,’ in  
198 which enzymes weakly catalyze non-canonical reactions in addition to their normal biological activities [35]. Under  
199 the right selection pressure, mutations can expand or improve these promiscuous activities, and evolution may  
200 eventually lead to specialization in the new activities [35]. By analogy, J might be said to have a baseline  
201 promiscuity in that it can weakly bind non-LamB-wt receptors, and that variants have increased promiscuity if  
202 they improve infectivity broadly over the space of potential hosts, rather than only on LamB-wt (Fig. 5c). To  
203 estimate the promiscuity of J variants, we plotted in log-space the progeny conferred by a variant on each host,  
204 compared to the susceptibility of these hosts to wild type  $\lambda$ . For synonymous variants (Fig. 5c, black line), this  
205 relationship is 1:1 by definition. However, some variants, like A84M, have a shallower slope indicating that they  
206 were less affected by the resistance of receptor mutations; we call variants with significantly greater area under  
207 the curve than synonymous variants ‘promiscuous.’ Most promiscuous variants grew well on the wild type host  
208 (Fig. S8), implying that these promiscuous variants could persist in a  $\lambda$  population prior to encountering a resistant  
209 host. However, that these variants have not already been fixed in the population may imply subtle costs to  
210 promiscuity, such as thermodynamic instability [35, 36].

211 Additionally, 49 single missense variants conferred growth on only a single host, most of which conferred  
212 growth on LamB-G267D (such as Q96E in Fig. 5d, S9). With a few exceptions, like K141C (Fig. 5d), these host-  
213 specific variants did not confer greater infectivity than wild type J on any host we tested; they were merely  
214 dramatically more infective on a particular host compared to the other hosts. Thus, while these host-specific  
215 variants may be an important part of adapting to a host, they are generally insufficient to directly overcome  
216 resistance. Rather, we posit that the acquisition of host-specific variants may follow after the acquisition of  
217 promiscuous variants and either ameliorate thermodynamic costs or prevent off-target binding. This process could  
218 explain why very few LamB-wt-specific variants were observed, and why they were less infective than variants  
219 specific to other hosts (Fig. S10), as if the wild type J sequence has already been selected for near-maximal  
220 specificity to its host.

221 Based on our analysis of J variants growing on a wild type host, positions with low mutational tolerance  
222 but high diversity over evolutionary time are predicted to be enriched within variants that drive adaptation to a  
223 novel receptor (Fig. 3c). Contrary to expectations, promiscuous variants tended to occur at positions with higher-



224 than-average progeny across variants (Fig. 5e). By contrast, host-specific variants tended to occur at positions that  
225 are intolerant to mutation and diversified over evolutionary time (Fig. 5f). Thus, while our hypothesis that these  
226 mutationally intolerant, evolutionarily diverse positions are driving host-specificity is largely supported, host-  
227 specificity is not equivalent to overcoming resistance, which can happen through host-nonspecific mechanisms  
228 (*i.e.*, promiscuity). This distinction implies that host range mutations arising in experimental evolution studies have  
229 not generally been promiscuous variants, as these host range mutations have mostly fallen at positions both  
230 mutationally intolerant and diverse (Fig. 3c). In our analysis, many of these host range mutations were null-like on  
231 all the hosts tested, including LamB-wt (Table S1), suggesting that the effects of these variants may be specific to  
232 the hosts used in those studies and/or co-occurring mutations in J.

233

### 234 **Positive epistasis potentiates adaptation to a new host**

235 In addition to single missense variants, the J library contained approximately 7,500 variants with two  
236 missense mutations, sparsely surveying the space of >4 million possible double missense variants. We wondered  
237 whether combinations of promiscuous and host-specific mutations could help mediate adaptation to a novel  
238 receptor beyond what either class of mutations would confer in isolation. For example, A94S is a promiscuous  
239 variant, but was mildly impaired for infectivity on LamB-G267D, and S30W is a LamB-G267D-specific variant that  
240 was mildly impaired on LamB-G267D but null-like on all others (Fig. 6a). However, the double missense variant  
241 A94S+S30W had both high growth and moderate specificity on LamB-G267D. This double missense variant  
242 therefore exhibits positive epistasis on one host, though it is poorly infective on the other hosts.

243 On LamB-wt, only 6.2% of double missense variants were infective (*i.e.*, not null-like), compared to 26%  
244 of single missense variants that were infective. For each double missense variant, we calculated the expected  
245 progeny given the progeny from each single variant, using a simple multiplicative model. Most variants yielded  
246 similar values for measured and expected progeny, but ~45% of infective variants (2.7% of total variants) exhibited  
247 significant epistasis (Fig. 6b), including 22 infective variants for which both mutations were null-like on their own.  
248 These variants with dramatic reciprocal sign epistasis were enriched for positions with host range mutations [3,  
249 14], which appeared 20 times among these 22 variants. As a control, we also considered variants in which a single  
250 missense mutation was paired with a synonymous mutation. Such variants exhibited better agreement between  
251 expected and empirical scores ( $r^2 = 0.92$ ) than double missense variants ( $r^2 = 0.64$ ), suggesting that the prevalence  
252 of epistasis in double missense variants is not an artifact of the fact that most double missense variants were less  
253 abundant in the library than single missense variants.

254 Similar to the selection on the wild type host, strong positive epistasis was prevalent on each resistant  
255 host. Single missense variants that conferred promiscuity frequently interacted epistatically with other variants

256 (Fig. 6c). Of double missense variants that contained a promiscuous single missense variant, 73/550 (13.3%)  
257 exhibited significant epistasis compared to only 2.2% of all double missense variants. The rare cases of double  
258 missense variants consisting of one promiscuous variant and one host-specific variant were even more likely to  
259 exhibit epistasis (5/21, 23.8%). In these five examples, epistasis was always in the positive direction.

260 Additionally, the effects of positive epistasis became more salient when  $\lambda$  was challenged with a resistant  
261 host. Double missense variants that were infective on LamB-wt had varying levels of promiscuity and positive  
262 epistasis (Fig. 6d). By contrast, double missense variants that were infective on the most resistant host, LamB-  
263 R219H, nearly exclusively had both high promiscuity and positive epistasis. The other two hosts revealed  
264 intermediate effects.

265

## 266 Discussion

267 Although phages and bacteria are often presumed to coevolve stably and indefinitely, an “asymmetry in  
268 coevolutionary potential of these hosts and parasites” exists [5]. To investigate how  $\lambda$  overcomes host resistance,  
269 we analyzed thousands of variants of its tail fiber protein, J, on a small set of resistant *E. coli* hosts. We find that  
270 promiscuous J variants, which increase infectivity on a broad range of hosts, underlie the re-establishment of  
271 infection on resistant hosts. These variants co-exist with other, host-specific, variants that generally do not  
272 increase infectivity on any host but have smaller losses to infectivity on a single host. We posit that both types of  
273 variants are important for a phage to adapt to a new host, with host-specific variants likely ameliorating costs  
274 associated with promiscuity. This framing has implications for experimental evolution studies, protein adaptation  
275 more broadly, and natural phage-bacteria communities.

276 When phages and bacteria cyclically develop resistance and counter-resistance in experimental evolution  
277 studies, this coevolution is frequently characterized by an initial escalation of both host resistance and phage  
278 counter-resistance. This escalation eventually reaches an asymptote and is followed by negative frequency-  
279 dependent selection (“Kill the winner” dynamics) in which the dominant phages are most infective on the most  
280 common hosts [6, 7]. This pattern is well explained by a model in which promiscuous variants drive broadened  
281 host range but come at a cumulative cost, manifesting as lower growth rate relative to their host-specific  
282 counterparts [37]. The sequential acquisition of promiscuous variants could also open up pathways for a phage to  
283 infect a highly resistant host by first adapting to a less resistant host in the same environment. For example, Werts  
284 *et al.* (1994) could not directly isolate  $\lambda$  that overcame the resistance allele LamB-G151D, but by pre-adapting the  
285 phage to other hosts with weaker resistance, they found double mutants able to grow on LamB-G151D [14].  
286 However, thermodynamic costs associated with promiscuity may also constrain paths to adaptation. In their  
287 isolation of LamB-independent  $\lambda$  strains that use OmpF as receptor, Meyer *et al.* (2012) repeatedly identified the

288 same set of 4-5 adaptive variants across dozens of independent cultures [3], suggesting considerable constraint  
289 on the path to counter-resistance [38]. In a follow-up study, the bi-specific intermediate, which binds both LamB  
290 and OmpF, was less stable than the LamB-specific parent, and selection for OmpF specificity was sufficient to  
291 restore stability [39].

292 In the context of enzymes, promiscuous activities provide convenient starting points for adaptation to  
293 novel protein functions [35]. Increasing the stability of the parent enzyme can potentiate greater adaptation by  
294 compensating for the mild destabilization associated with some adaptive variants [36, 40]. Similarly, the low level  
295 of infectivity mediated by wild type J on resistant hosts serves as a starting point for adaptation to those hosts.  
296 Second mutations can compensate for the initial mutations that confer promiscuity, offsetting the potential costs  
297 of these initial mutations and resulting in highly infective and/or promiscuous double variants. This effect can be  
298 seen in the strong association between positive epistasis and promiscuity, whereby promiscuous single variants  
299 were more likely to have positive epistatic interactions than non-promiscuous variants (Fig. 6c) and rare  
300 promiscuous variants that drive adaptation to LamB-R219H were always positively epistatic (Fig. 6d). At a  
301 mechanistic level, promiscuous J variants may shift between multiple semi-stable protein conformations, or they  
302 may heterogeneously fold into one of multiple stable conformations. The latter mechanism was found to underlie  
303 the LamB-OmpF bi-specific intermediates characterized by Petrie *et al.* (2018) [39]. Under a model requiring  
304 multiple protein conformations, destabilization may be fundamentally linked to promiscuity rather than incidental  
305 to it. Therefore, counter to observations with promiscuous enzymes [36, 40], stabilizing mutations in phages are  
306 unlikely to precede adaptive ones. Instead, a destabilizing mutation that increases promiscuity must come first,  
307 followed by a compensatory mutation that re-stabilizes the protein into an optimal conformation for infection of  
308 the most abundant host. This prediction would also explain why J is so broadly intolerant of mutation: repeated  
309 evolutionary transitions between stable and unstable sequences leave J close to a threshold of severe  
310 destabilization, compared to proteins whose evolutionary histories are dominated by stable sequences [41, 42].

311 Naturally occurring phage–bacteria interactions also show patterns consistent with a balance between  
312 host-specific and promiscuous variation. Phages and bacteria isolated from the same environment form infection  
313 networks exhibiting both nestedness and ‘modularity,’ a property of lock-key models, with nestedness dominating  
314 at small scales involving highly related strains [43, 44]. This pattern is consistent with promiscuity driving counter-  
315 resistance to a newly resistant host, and modularity arising between more diverged hosts. The extent to which  
316 the evolution of phages involves adaptation to diverged hosts remains unclear. Orthologs of both J and LamB are  
317 broadly distributed among enterobacteria, and even appear in distant  $\epsilon$ -proteobacteria species, suggesting either  
318 an ancient origin of this host–pathogen relationship or frequent cross-taxa jumps in host range. However, the  
319 limited breadth of hosts in which nestedness is observed suggests practical constraints to the promiscuity of a

320 phage tail: a single promiscuous phage variant is more likely to be infective multiple receptor variants within a  
321 single host species than on receptors of multiple related host species. In our hands, an *E. coli* host expressing a  
322 LamB ortholog from another enterobacteria (*C. freundii*, *Y. pestis*, or *S. marcescens*) did not support growth of our  
323 library, which mostly contained single missense mutations, suggesting that multiple mutations may be required  
324 for jumps between species.

325 We conclude that although  $\lambda$  faces significant evolutionary hurdles not faced by its host, it can establish  
326 common paths to adaptation on multiple potential hosts by maintaining a balance between promiscuity and host-  
327 specificity. This balance may be mediated by mild destabilization of the protein, allowing it to sample multiple  
328 conformations, although further work is needed to directly test this hypothesis. Although we surveyed adaptation  
329 to only a small set of resistant hosts, this general framework is consistent with prior observations of how  $\lambda$  evolves  
330 to switch to a novel receptor. This framework may apply broadly to other phages and viruses for which mutations  
331 are more difficult to assay *en masse* [7, 45, 46].

332

### 333 **Acknowledgements**

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

337

### 338 **Ethics Declarations**

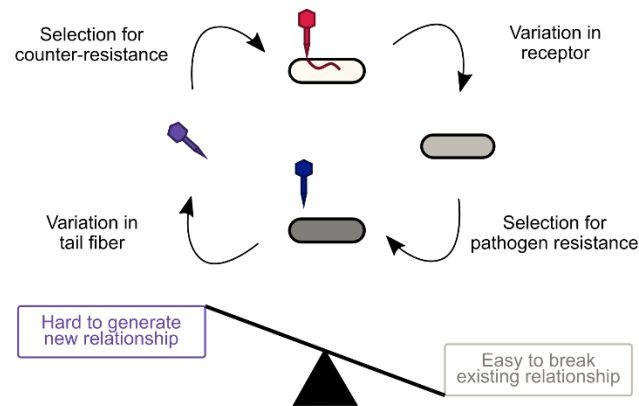
339 Conflicts of Interest: The authors declare that they have no conflicts of interest.

## 340 References

- 1 Morona R, Henning U. Host Range Mutants of Bacteriophage Ox2 Can Use Two Different Outer Membrane Proteins of Escherichia coli K-12 as Receptors. *J Bacteriol.* 1984;159:579–82.
- 2 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.
- 3 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.
- 4 Lenski RE. Coevolution of bacteria and phage: Are there endless cycles of bacterial defenses and phage counterdefenses? *J Theor Biol.* 1984;108:319–25.
- 5 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.
- 6 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.
- 7 Hall JPJ, Harrison E, Brockhurst MA. Viral host-adaptation: Insights from evolution experiments with phages. *Curr Opin Virol.* 2013;3:572–7.
- 8 Dennehy JJ. What Can Phages Tell Us about Host-Pathogen Coevolution? *Int J Evol Biol.* 2012;2012.
- 9 Buckling A, Rainey PB. Antagonistic coevolution between a bacterium and a bacteriophage. *Proc R Soc B Biol Sci.* 2002;269:931–6.
- 10 Lederberg E. Pleiotropy for maltose fermentation and phage resistance in Escherichia coli K-12. *Genetics.* 1955;40:580–1.
- 11 Thirion JP, Hofnung M. On some genetic aspects of phage  $\lambda$  resistance. *Genetics.* 1972;71:207–16.
- 12 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.
- 13 Shaw JE, Bingham H, Fuerst CR, Pearson ML. The Multisite Character of Host-Range Mutations in Bacteriophage  $\lambda$ . *Virology.* 1977;83:180–94.
- 14 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.
- 15 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.
- 16 Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, et al. Antagonistic coevolution accelerates molecular evolution. *Nature.* 2010;464:275–8.
- 17 Gurnev PA, Oppenheim AB, Winterhalter M, Bezrukov SM. Docking of a Single Phage Lambda to its Membrane Receptor Maltoporin as a Time-resolved Event. *J Mol Biol.* 2006;359:1447–55.
- 18 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.
- 19 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.
- 20 Hyman P, van Raaij M. Bacteriophage T4 long tail fiber domains. *Biophys Rev.* 2018;10:463–71.
- 25 Fowler DM, Fields S. Deep mutational scanning: A new style of protein science. *Nat Methods.* 2014;11:801–7.
- 21 Wrenbeck EE, Klesmith JR, Stapleton JA, Adeniran A, Tyo KEJ, Whitehead TA. Plasmid-based one-pot saturation mutagenesis. *Nat Methods.* 2016;13:928–30.
- 22 Kaiser D, Masuda T. In vitro assembly of bacteriophage Lambda heads. *PNAS.* 1973;70:260–4.
- 23 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.
- 24 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.
- 26 Roscoe BP, Bolon DNA. Systematic exploration of ubiquitin sequence, E1 activation efficiency, and experimental fitness in yeast. *J Mol Biol.* 2014;426:2854–70.

- 27 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.
- 28 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.
- 29 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.
- 30 Boucher JJ, 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.
- 31 Andrews B, Fields S. Distinct patterns of mutational sensitivity for  $\lambda$  resistance and maltodextrin transport in *Escherichia coli* LamB. *Microb Genomics.* 2020;6.
- 32 Heine HG, Francis G, Lee KS, Ferenci T. Genetic analysis of sequences in maltoporin that contribute to binding domains and pore structure. *J Bacteriol.* 1988;170(4):1730–8.
- 33 Schirmer T, Keller TA, Wang Y, Rosenbusch JP. Structural Basis for Sugar Translocation Through Maltoporin Channels at 3.1 Å Resolution. *Science.* 1995;267:512–4.
- 34 Weitz JS, Hartman H, Levin SA. Coevolutionary arms races between bacteria and Bacteriophage. *Proc Natl Acad Sci U S A.* 2005;102:9535–40.
- 35 Arnold FH. How proteins adapt: Lessons from directed evolution. *Cold Spring Harb Symp Quant Biol.* 2009;74:41–6.
- 36 Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. *PNAS.* 2006;103:5869–74.
- 37 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.
- 38 Franke J, Kloser A, Visser JAGM De, Krug J. Evolutionary Accessibility of Mutational Pathways. *PLoS Comput Biol.* 2011;7.
- 39 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.
- 40 Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol.* 2009;19:596–604.
- 41 Gong LI, Suchard MA, Bloom JD. Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife.* 2013;2:e00631.
- 42 Tokuriki N, Tawfik DS. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature.* 2009;459:668–73.
- 43 Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS. Statistical structure of host-phage interactions. *PNAS.* 2011;108:E288–97.
- 44 Weitz JS, Poisot T, Meyer JR, Flores CO, Valverde S, Sullivan MB, et al. Phage-bacteria infection networks. *Trends Microbiol.* 2013;21:82–91.
- 45 Carrat F, Flahault A. Influenza vaccine: The challenge of antigenic drift. *Vaccine.* 2007;25:6852–62.
- 46 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.
- 47 Rajagopala S V, Casjens S, Uetz P. The protein interaction map of bacteriophage lambda. *BMC Microbiol.* 2011;11.
- 48 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.

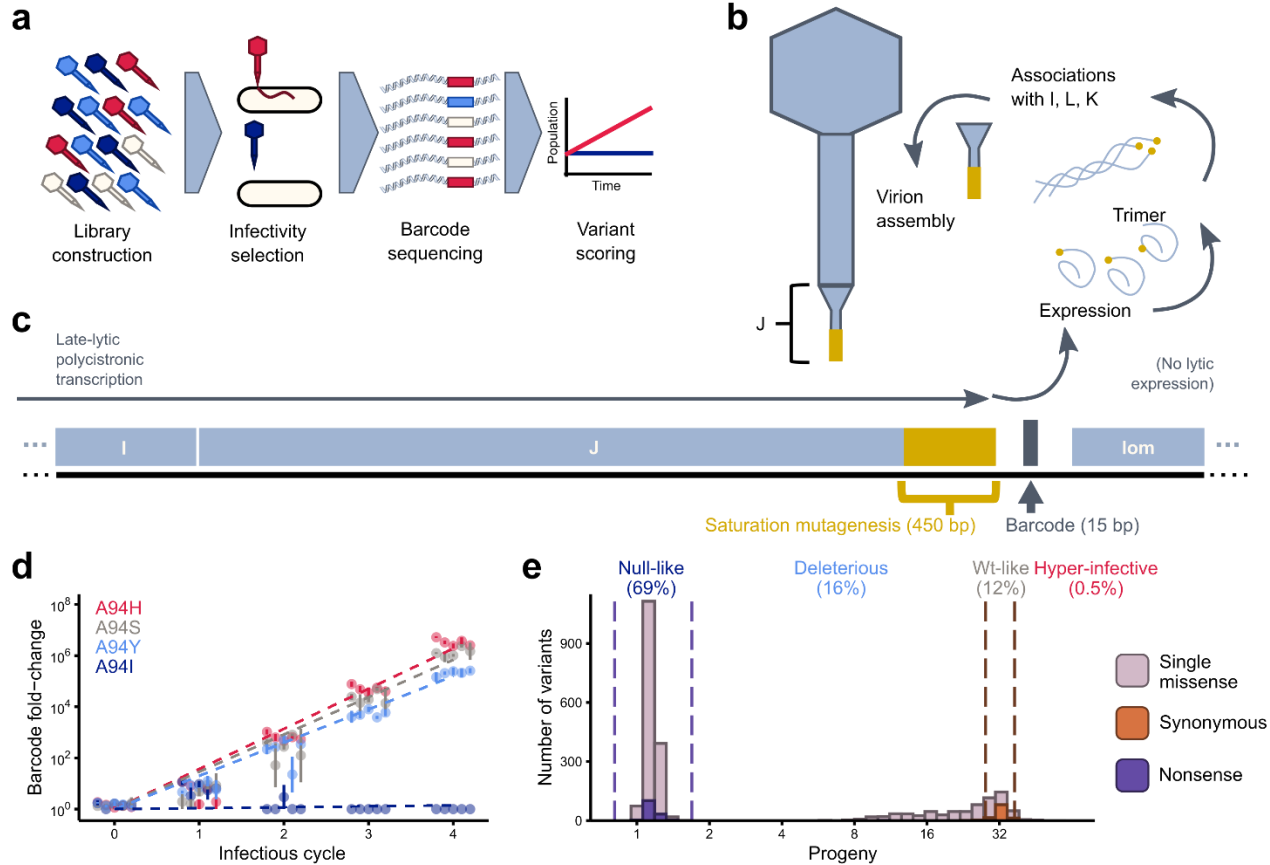
342 **Figures**



343

344 **Figure 1: Phage–bacteria coevolution is fundamentally weighted in favor of bacteria.** Phages and bacteria are presumed to  
345 coevolve stably and indefinitely in the wild, but theoretical considerations and experimental co-culture systems mostly  
346 predict long-term dominance of bacteria. Explaining the success of phages over evolutionary timescales is further hampered  
347 the limited mechanistic understanding of how phages adapt to a new and/or resistant host.  
348

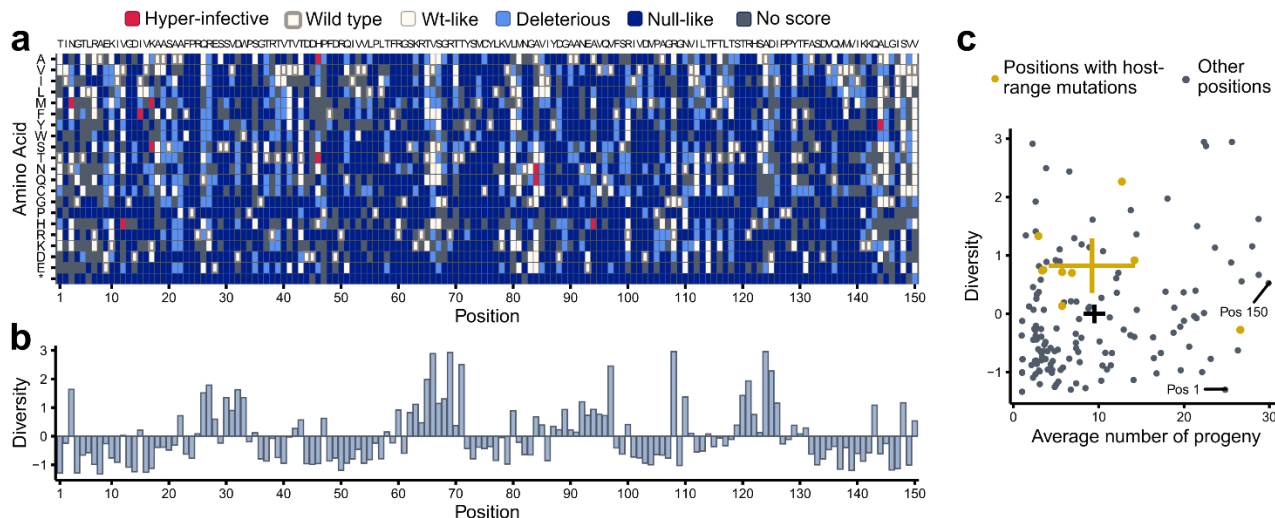
349  
350



351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368

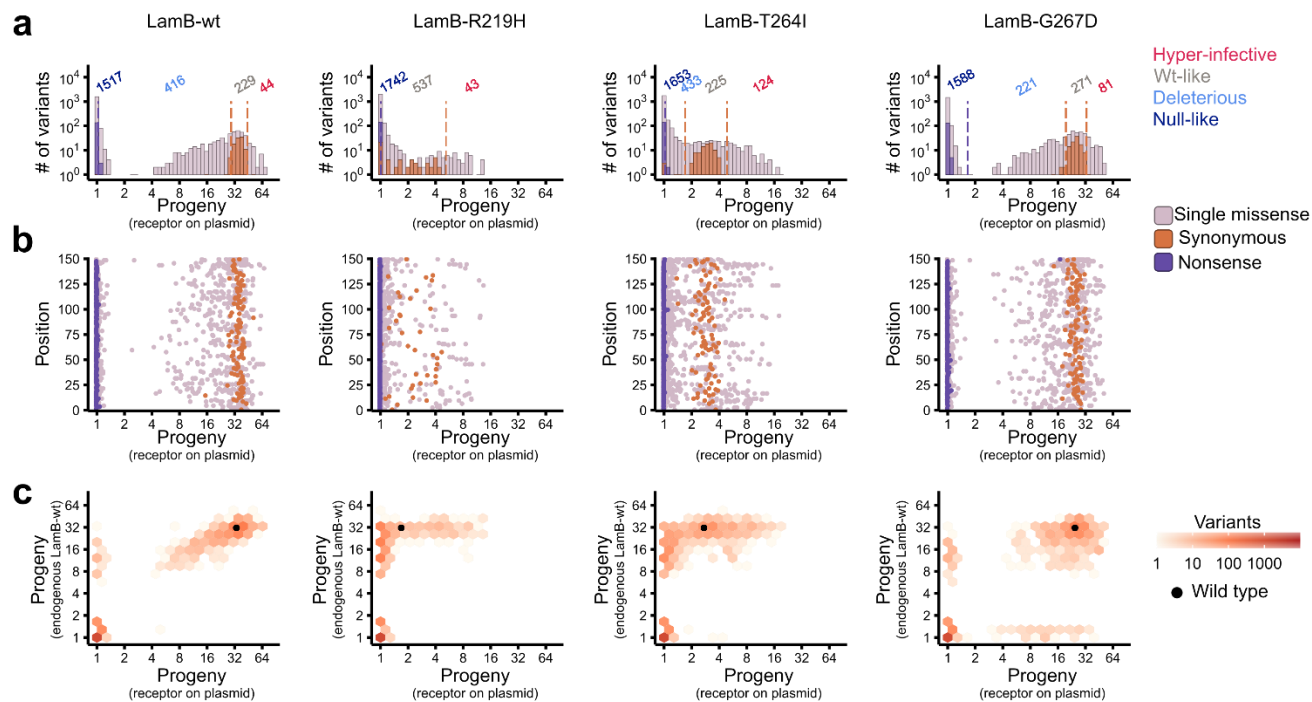
**Figure 2: A sequencing-based method to assess phage infectivity across thousands of variants.** (a) The experimental strategy. A library of phage variants was constructed and subjected to selection for the ability to infect *E. coli*. After each infectious cycle, barcodes corresponding to each variant 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 [47]. (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 3 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. Dashed lines indicate score boundaries used to categorize variants as null-like, deleterious, wt-like, or hyper-infective.



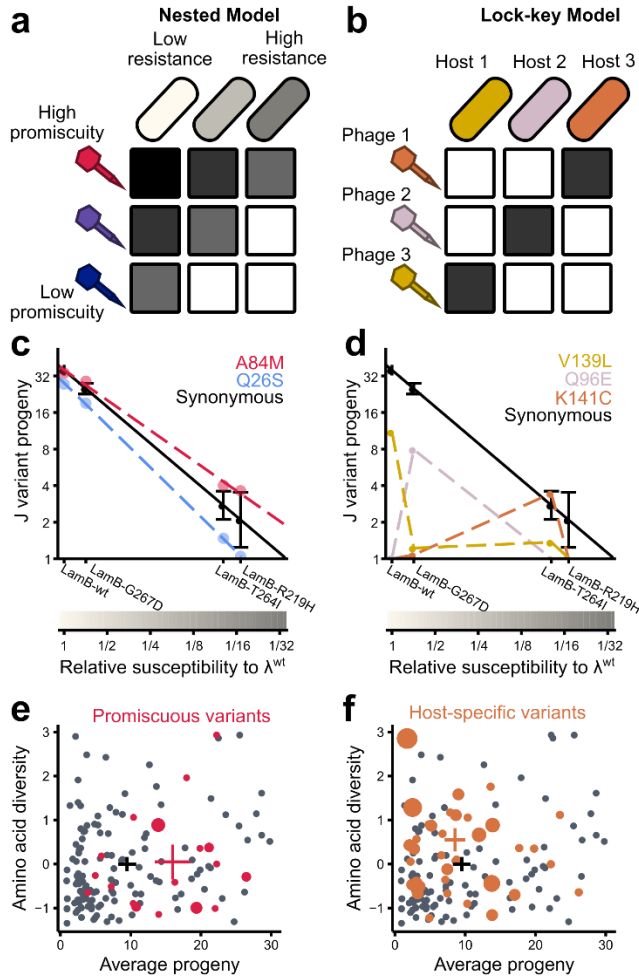


369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380

**Figure 3: Comparison of empirical mutational tolerance and evolutionary diversity.** (a) Categorical effects of all amino acid substitutions. Categories are defined in Fig. 2e. 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 [48], 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 [26, 27, 30]. Gold points indicate positions where mutations have been reported that expand host range [3, 14]; 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.



381  
 382 **Figure 4: Selection of  $\lambda$  bearing J variants on  $\lambda$ -resistant hosts. (a)** Distribution of fitness effects on each host. In each case,  
 383 the synonymous and nonsense distributions can be separated, despite the nominal  $\lambda$ -resistance of each lamB allele. Variants  
 384 that outperform wild type  $\lambda$  do so by a larger margin on hosts that are more resistant. **(b)** Progeny for each J variant shown  
 385 by the position of the mutation in the sequence. Adaptive mutations occur frequently at a subset of positions, but these  
 386 positions are spread over the entire mutagenized region. Positions with many hyper-infective mutations are more apparent  
 387 on the more resistant hosts LamB-R219H and LamB-T264I than on LamB-wt or LamB-G267D. **(c)** Correlation between progeny  
 388 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).  
 389 Variants that are highly infective on a non-wt host tend to also be infective on the wild type host, with some exceptions.  
 390 However, many variants that are highly infective on the wild type host are poorly infective on non-wt hosts.  
 391



392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

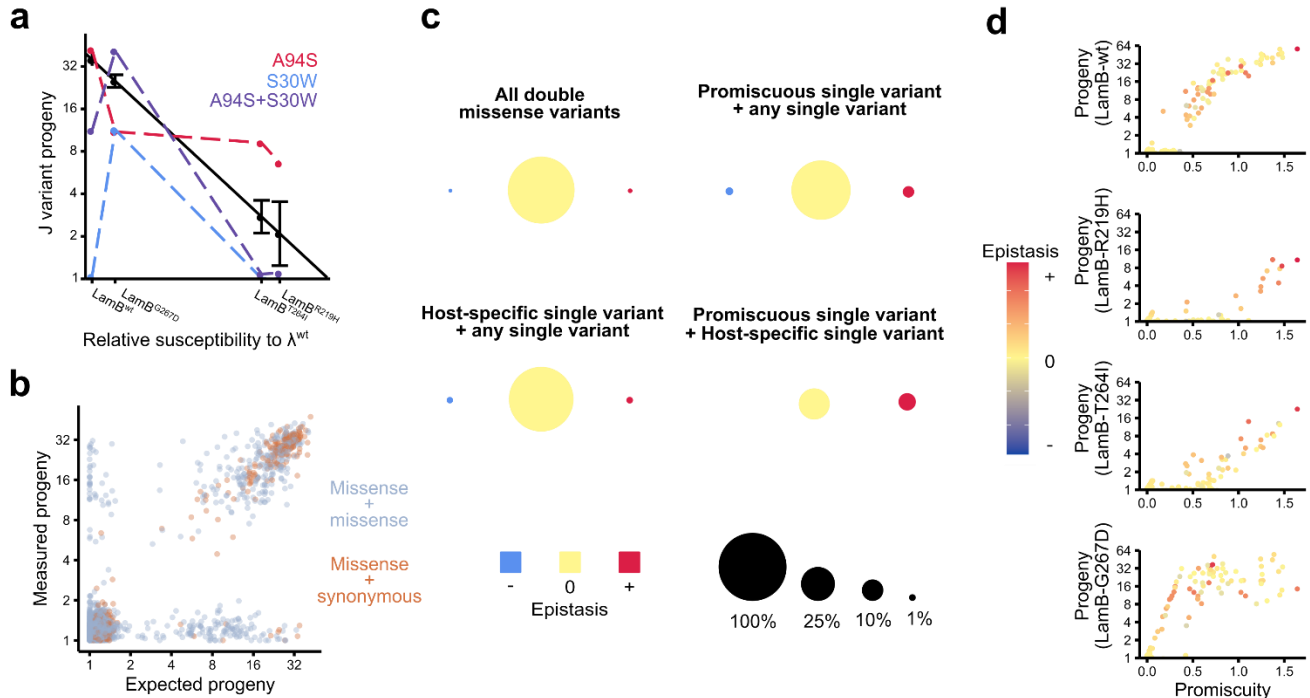
407

408

409

410

**Figure 5: 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. We calculate promiscuity as the area under the curve for a regression line comparing the variant infectivity to host susceptibility, relative to wild type  $\lambda$ . The black line represents synonymous variants, with error bars equal to  $\pm 1$  standard deviation. **(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.



411

412

**Figure 6: Double missense variants in J can mediate adaptation to new hosts. (a)** The double missense variant S30W, A94S

413

is a combination of a promiscuous variant (A94S) and a LamB-G267D-specific variant (S30W). The double missense variant

414

exhibits sign epistasis, improving infectivity on LamB-G267D despite the deleteriousness of each single missense variant. **(b)**

415

For double missense variants (silver), we calculated the expected progeny from the progeny of each of the single missense

416

variants using a simple multiplicative model. A subset of double missense variants strongly deviates from the multiplicative

417

model, in contrast to variants in which a single missense mutation is paired with a single synonymous mutation (orange),

418

which are more likely to agree with the multiplicative model ( $r^2 = 0.92$  missense x synonymous vs.  $r^2 = 0.64$  missense x

419

missense). **(c)** Across the four hosts, most double missense variants in J do not exhibit significant epistasis (top left panel).

420

However, double missense variants that contain a promiscuous variant, a host-specific variant, or both, are more likely to

421

exhibit significant epistasis. We measured significant epistasis in 13.2% of double missense variants containing promiscuous

422

variants, 5.4% containing host-specific variants, and 23.8% containing both, compared to 1.8% containing neither. **(d)**

423

For each host, progeny is positively associated with promiscuity and with positive epistasis. However, these associations become

424

more salient on resistant hosts, with all infective variants on LamB-R219H being promiscuous and positively epistatic,

425

compared to a minority of variants on LamB-wt.