DOI: 10.1534/genetics.104.039438

The Rate of Compensatory Mutation in the DNA Bacteriophage ϕX174

Art Poon[1] **and Lin Chao**

Division of Biological Sciences, Section of Ecology, Behavior and Evolution, University of California, San Diego, California 92093-0116

Manuscript received December 7, 2004
Accepted for publication March 28, 2005

ABSTRACT

A compensatory mutation occurs when the fitness loss caused by one mutation is remedied by its epistatic interaction with a second mutation at a different site in the genome. This poorly understood biological phenomenon has important implications, not only for the evolutionary consequences of mutation, but also for the genetic complexity of adaptation. We have carried out the first direct experimental measurement of the average rate of compensatory mutation. An arbitrary selection of 21 missense substitutions with deleterious effects on fitness was introduced by site-directed mutagenesis into the bacteriophage ϕX174. For each deleterious mutation, we evolved 8–16 replicate populations to determine the frequency at which a compensatory mutation, instead of the back mutation, was acquired to recover fitness. The overall frequency of compensatory mutation was ~70%. Deleterious mutations that were more severe were significantly more likely to be compensated for. Furthermore, experimental reversion of deleterious mutations revealed that compensatory mutations have deleterious effects in a wild-type background. A large diversity of intragenic compensatory mutations was identified from sequencing fitness-recovering genotypes. Subsequent analyses of intragenic mutation diversity revealed a significant degree of clustering around the deleterious mutation in the linear sequence and also within folded protein structures. Moreover, a likelihood analysis of mutation diversity predicts that, on average, a deleterious mutation can be compensated by about nine different intragenic compensatory mutations. We estimate that about half of all compensatory mutations are located extragenically in this organism.

COMPENSATORY mutation is an important and poorly understood evolutionary process. A compensatory mutation occurs when the loss of fitness caused by one mutation is remedied by its epistatic interaction with a second mutation at another site in the genome (Kimura 1985). Are compensatory mutations common, or do they represent a genetic oddity? We require some empirical estimate of the average rate or number of compensatory mutations, not only because it would inform us about the evolutionary consequences of mutation accumulation (Whitlock and Otto 1999), but also because it would quantify the genetic complexity of adaptation (Orr and Coyne 1992). Consider, for example, the ratio of the rate of compensatory mutation and the back mutation rate. This ratio determines the probability that a population that is fixed for a deleterious mutation will recover by acquiring some alternative to the ancestral genotype. To illustrate, we refer to the evolution of resistance in microbial pathogens. Mutations that confer resistance to an antimicrobial agent are often detrimental to growth in the drug-free environment (Maisnier-Patin and Andersson 2004). The probability that a population in this environment acquires a compensatory mutation and retains resistance, instead of reverting the resistance-conferring mutation, is dependent on the ratio of mutation rates (Levin *et al.* 2000). To date, no experimental work directly measures the average rate of compensatory mutation. Although an abundance of studies characterizes suppressor mutations (Sujatha and Chatterji 2000), which prevent the phenotypic expression of some prior mutation, most cases not only lack an unambiguous association with fitness evolution, but also are limited to the suppression of only one or two mutations. However, a recent likelihood analysis of the suppressor mutation literature predicts that a deleterious mutation is associated on average with >10 potential sites for compensatory mutation (Poon *et al.* 2005, this issue), regardless of whether the organism is a virus, a prokaryote, or a eukaryote. This result implies that a meaningful average number of compensatory mutations can be isolated from an arbitrary group of deleterious mutations within any model organism.

To measure the average rate of compensatory mutation, we have generated a large number of deleterious mutations derived from a high-fitness ancestral genotype of the DNA bacteriophage ϕX174. This bacteriophage has a compact circular genome that consists of 5386 nucleotides coding for 11 genes (designated by the letters *A–H, J, K,* and *A** for a gene nested within A; Hayashi *et al.* 1988). A number of suppressor mutations

[1]*Corresponding author:* Division of Biological Sciences, University of California, 9500 Gilman Dr., Muir Biology Bldg., Room 3155, San Diego, CA 92093-0116. E-mail: apoon@biomail.ucsd.edu

affecting several gene products in ϕX174 have previously been characterized (*e.g.*, Fane and Hayashi 1991), implying that epistatic interactions that can produce compensatory mutations are not uncommon in this organism. Moreover, the structure of the ϕX174 polypeptide procapsid assembled from the products of genes *F*, *G*, and *D* is well characterized (McKenna *et al.* 1992; Dokland *et al.* 1997) and provides a basis for characterizing functional interactions between deleterious and compensatory mutations. In our experiment, we estimated the ratio of compensatory mutation to back mutation by determining the frequency at which replicate populations recovered fitness while retaining the deleterious mutation. Furthermore, we identified intragenic compensatory mutations by sequencing from each replicate an interval of the genome containing the gene affected by the deleterious mutation.

Our experiment creates a unique opportunity to estimate the average rate of compensatory mutation for a large arbitrary selection of deleterious mutations within a single ancestral genotype. In addition, it has also yielded the following novel results: (1) that there is a significant positive association between the severity of a deleterious mutation and the rate of compensatory mutation; (2) that compensatory mutations overall are significantly clustered around deleterious mutations in the linear amino acid sequence and also exhibit structural clustering in folded proteins; and (3) that the average proportion of compensatory mutations that are located extragenically is ~50%. These results have important implications for our understanding of adaptation, the evolution of resistance, and the evolutionary consequences of biological complexity in viruses and higher organisms.

MATERIALS AND METHODS

Construction of deleterious mutations: A wild-type strain of the ϕX174 bacteriophage and its host *Escherichia coli C* (*C122*) were generously provided by B. A. Fane. This bacteriophage strain differs from the Sanger *et al.* (1978) genome sequence (accession no. NC001422) at 11 nucleotide sites, causing one silent (C2811T), one nonsense (C367T), and nine missense (G833A, C1460G, A1650G, C1727T, C2085T, A2731G, C3121A, G3139A, and G4518A) substitutions. From this wild-type strain, we derived two populations that were evolved by serial transfer of $\sim 10^4$ phage on petri plates supplemented with solid LC media (modified Luria broth with 1.5% w/v agarose; Chao *et al.* 1997). All plates were incubated at 33° for 4-hr intervals (*e.g.*, growth cycle), each having a layer of 3 ml top agar (LC with 0.7% w/v agarose) inoculated with 0.3 ml from an overnight culture of *C122* and supplemented with 10 mm $MgCl_2$ and 5 mm $CaCl_2$ to promote infection. After growth, the top agar was harvested into 3 ml BE buffer (50 mm $Na_2B_4O_7$, 3.0 mm EDTA; Fane and Hayashi 1991) and three drops of $CHCl_3$ to terminate bacterial growth and elute phage from cellular debris. After 300 growth cycles, we isolated a high-fitness clone (here referred to as 300*aI*) from one population to serve as a common ancestral genotype from which all deleterious mutations would be derived for this experiment. The evolved 300*aI* genome sequence differs from the wild-type strain at seven nucleotide sites, causing one silent (T1345C) and six missense (A1301G, A1428G, C3130G, A3189G, A3121G, and G3339A) substitutions.

Our next objective was to generate a large number of mutations in 300*aI* that were engineered to be easily detectable. We located all nucleotide substitutions in the 300*aI* genome that would introduce a GATC restriction site, recognized by the *Dpn*II restriction enzyme (New England Biolabs, Beverly, MA). Because there are no naturally occurring GATC sites in the ϕX174 genome, every such substitution creates a unique restriction site. We screened these substitutions for a subset that would result in a missense substitution in a protein-coding region of the genome. Additionally, three substitutions (C1137T, C1280T, and C2480T) that do not introduce a *Dpn*II restriction site were included in this set of targets. For every target substitution, we designed and ordered a site-specific 18- to 23-mer primer (Integrated DNA Technologies) with a mismatched nucleotide corresponding to a target substitution in the ancestral 300*aI* sequence. We used a PCR-based method of site-directed mutagenesis by amplifying a 1- to 2-kbp segment from a phenol:chloroform extraction of the 300*aI* genome using a mismatched primer complemented by a second primer. The entire reaction mixture was subsequently transformed by heat shock into competent *C122* cells prepared by a cold $CaCl_2$ treatment (Sambrook *et al.* 1989). The partial mutant genome is spontaneously completed within the *C122* cells, possibly by cell-mediated extension of the lagging strand over the wild-type template and ligation to form a double-stranded circular genome that is heterozygous for the target mutation. The transformation mixture was plated with an inoculum of *C122* and incubated for 4 hr at 33°. Emergent plaques were isolated and stored at −20° in a 1:1 mixture of 80% w/v glycerol and HFB buffer [0.1 m Tris-HCl (pH 7.4), 0.06 m NH_4Cl, 0.09 m NaCl, 0.1 m KCl, 1 mm $MgSO_4$, 1 mm $CaCl_2$; Fane and Hayashi 1991]. Plaque isolate genotypes were identified by a *Dpn*II restriction digest of an amplified segment containing the target nucleotide (excepting the three targets mentioned above, which were confirmed by sequencing). Isolates that, following restriction digest, produced a multiple band phenotype in an electrophoretic gel were evaluated for a reduction of fitness by a plaque size assay (described below). If the mutant plaque isolate exhibited a deleterious effect on fitness, then the isolated genome was sequenced completely to confirm that the target nucleotide was the only site in the genome that differed from the ancestral genotype 300*aI*.

We generated 21 different deleterious mutations by this method. To confirm that each substitution introduced by site-directed mutagenesis was responsible for the fitness reduction, we reverted that nucleotide by the same procedure unless a back mutation had been already acquired from serial transfers of the deleterious mutation. In addition to this set of deleterious mutations, we also studied two mutations (G2979A and A5031G) that occurred spontaneously during site-directed mutagenesis. Although these mutations were accompanied by other substitutions in the genome, reversion of the latter had no effect on the fitness reduction in these mutant genotypes. Reversion of the mutations G2979A and A5031G restores fitness. The complete set of deleterious mutations used in this study is listed in Table 1.

Measurement of fitness: We used a plaque size assay to estimate the fitness of mutant genotypes. The average plaque area was measured from digital images of replicate plates of a given genotype, each containing a standardized volume of solid LC media (30 ml). Digital images were captured using a Cohu 4900 series high-performance CCD camera and analyzed using Scion Image version 3b (Scion, Frederick, MD). Plaque size has previously been shown to be strongly correlated with fitness

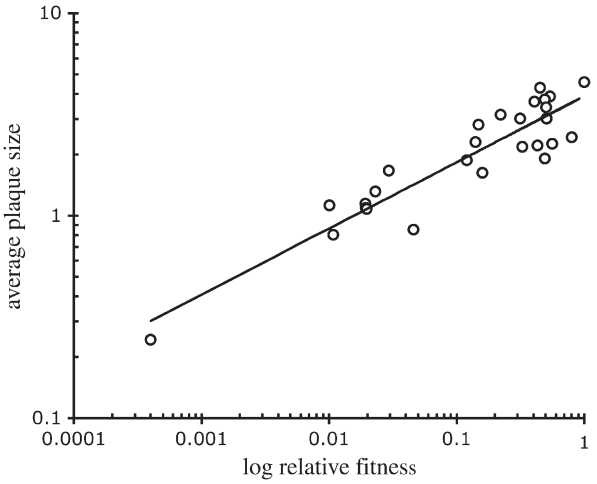

FIGURE 1.—Log-log plot of average plaque size against relative fitness. Average plaque size is in square millimeters. Relative fitness is estimated by the average yield per plaque normalized by the ancestor 300*aI*. The top-rightmost data point represents the 300*aI* high-fitness ancestor. A solid line is drawn to illustrate a linear fit to the log-transformed data ($R^2 = 0.85$). The two fitness measures are strongly correlated (Pearson's $r = 0.92$; 95% C.I. = 0.83–0.96; $P < 10^{-5}$), which thereby implies that plaque size is a justifiable proximate measure of fitness. In generating this plot, we have included fitness measures from two additional genotypes not used in this study.

in experimental populations of the RNA bacteriophage ϕ6 (BURCH and CHAO 2004). To determine the reliability of plaque size as a measure of fitness in ϕX174, we measured both the average plaque size and the average yield of phage per plaque for all 23 deleterious mutations and the 300*aI* ancestor and evaluated the correlation between the two fitness measures. Average yield of phage per plaque was determined by measuring the phage density in lysate recovered from replicate plates containing a known number of plaques. For both plaque size and plaque yield assays, we plated ~100 plaques to minimize plaque overlap. We found a strong correlation between log-transformed values of average plaque size and average yield (Pearson's $r = 0.92$; Figure 1). This correlation justifies the use of plaque size as a proxy for fitness in ϕX174.

Screening for compensatory mutations: For each deleterious mutation, we initiated 8–16 replicate populations that were subsequently evolved by the serial transfer of 10^3–10^4 phage until an increase in plaque size was detected. A plaque size increase was detected either by visual inspection or by the computer-assisted measurement of average plaque size. Upon detecting a significant increase, we isolated two plaques from that population and verified the divergence of plaque size from the original deleterious mutant. Hence, we did not require the recovered genotype to fix in the population. All plaque isolates were stored individually at −20° in a 1:1 glycerol:HFB mixture. In most cases, we used *Dpn*II restriction to determine whether the deleterious mutation introduced by site-directed mutagenesis was retained in the plaque isolates that had recovered fitness. Otherwise, an interval containing the site of deleterious mutation was sequenced. Retention of the deleterious mutation implies that the recovery of fitness was caused by a compensatory mutation at another site in the genome. For every replicate population potentially containing a compensatory mutation, we sequenced an interval of the genome surrounding the gene in which we had introduced a deleterious mutation. Any substitution observed in that interval was recorded as a putative intragenic compensatory mutation. (No more than one new substitution was observed in any sequence from compensated genotypes with deleterious mutations generated by site-directed mutagenesis.) An absence of substitutions in the gene containing the deleterious mutation was interpreted as a case of extragenic compensatory mutation. The positions of deleterious and compensatory mutations in the protein structure were visualized using the University of California at San Francisco (UCSF) Chimera v1.2 package (Computer Graphics Laboratory, UCSF; PETTERSEN *et al.* 2004). We obtained a published structure of the ϕX174 procapsid subunit (PDB ID: 1AL0; DOKLAND *et al.* 1997) from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (BERMAN *et al.* 2000).

All statistical and randomization analyses were performed in the *R* statistical environment for Macintosh (R Development Core Team). The likelihood analysis of intragenic mutation diversity followed the procedure described in previous work (POON *et al.* 2005, this issue).

RESULTS

The frequency of compensatory mutation: The numbers of replicate populations that acquire either back mutations or compensatory mutations are listed in Table 1 for replicate populations descended from each of the 23 deleterious mutations in our experiment. The overall frequency at which a replicate population acquired a compensatory mutation instead of a back mutation was equal to 67.7%, calculated as a weighted average across all deleterious mutations (Table 1). For two (A3582G and G4970T) of the 23 deleterious mutations, none of the populations recovered fitness after 50 growth cycles by serial transfer. Because A3582G and G4970T have weak effects on both the yield of phage per plaque and the average plaque size (Table 1), any back or compensatory mutations not only would be very slow to increase in frequency, but also would have to reach a substantial frequency to cause a detectable increase in average plaque size. Both mutations were omitted from further analyses in our study. Of the 21 remaining deleterious mutations, there are 4 mutations (T489C, T495C, C1280T, and C4552A) for which only back mutations were isolated in all replicate populations (Table 1), implying that these particular mutations cannot be compensated by mutations at any other site in the genome. Two of these mutations (T489C and T495C) both occur in gene *D*, which encodes the external scaffolding protein of ϕX174 assembly and is the most conserved gene in the family Microviridae (BURCH and FANE 2000). Conversely, no back mutations were observed in any fitness-recovered populations fixed for one of four other deleterious mutations (G232C, C2480T, A3054C, and G3524C; Table 1). For these mutations, the lack of back mutations implies that the rate of compensatory mutation greatly exceeds the back mutation rate.

It is important at this point to make a distinction between the actual number of compensatory mutations and the evolutionary contribution of compensatory mu-

TABLE 1

Summary of deleterious mutations

Mutation	Gene	Amino	Plaque size	Relative fitness	n	n_{back}	n_{comp}
G232C	*C*	E34Q	0.803	0.0108	16	0	14
T489C	*D*	F34L	2.446	0.7946	16	16	0
T495C	*D*	F36L	2.270	0.5609	16	16	0
G1132C	*F*	M43L	2.818	0.1485	16	3	13
C1137T	*F*	A45V	3.020	0.3151	16	3	13
C1280T	*F*	P93S	3.659	0.4081	16	3	0
G1423C	*F*	M140I	2.186	0.3257	16	5	11
A1624T	*F*	E207D	1.092	0.0195	16	3[a]	12
T1631C	*F*	Y210H	4.279	0.4509	8	2	2
A1757T	*F*	T252S	1.671	0.0295	8	1	7
C2235T	*F*	T411I	3.161	0.2214	8	1	5
C2480T	*G*	A29V	1.881	0.1205	16	0	12
G2746C	*G*	G118R	0.244	0.0004	8	0	8
G2979A[b]	*H*	A17T	2.222	0.4283	8	5	3
A3054C	*H*	N42H	1.920	0.4901	16	0	16
T3288C	*H*	Y120H	2.317	0.1405	8	4	4
G3326T	*H*	E132D	1.124	0.0101	16	2	14
G3524C	*H*	M198I	3.021	0.5071	16	0	16
A3582G	*H*	N218D	3.873	0.5385	16	0	0
C4552A	*A*	A191D	1.632	0.1589	8	8	0
G4970T	*A*	E330D	3.743	0.4895	16	0	0
A5031G[b]	*A*	K351E	1.080	0.0197	16	8	8
C5337A	*A, B*	L453I, A88D	1.318	0.0230	16	2	14
300*aI*			4.564	1.0			
					Totals: 312	82	172

Each mutation is identified by the initial nucleotide, genome position (numbered from the *Pst*I restriction site in accordance with Sanger *et al.* 1978), and final nucleotide. "Amino" refers to amino acid substitutions (initial amino acid, position, and final amino acid). "Plaque size" (*i.e.*, area) is reported in units of millimeters squared. "Relative fitness" is estimated from the average phage yield per plaque of the deleterious mutation, standardized by the yield of ancestral 300*aI*. The total number of replicate populations founded from each deleterious mutation is listed under n. The numbers of populations acquiring back or compensatory mutations are listed under n_{back} and n_{comp}, respectively.

[a]Includes two independent same-site reversions to nonancestral nucleotide (T1624G).

[b]Does not introduce *Dpn*II restriction site.

tation. Not every possible compensatory mutation contributes equally to the evolution of fitness. Either a diminishingly small effect on fitness or a relatively low mutation rate can prevent a compensatory mutation from becoming established in a modest-sized population. Ideally, we want to detect every possible compensatory mutation. However, compensatory mutations that are difficult to observe because of their mutation rate or fitness effect require a large population size, in which case the back mutation is ensured to out-compete them (Burch and Chao 1999). Thus, the frequency of compensatory mutation in our experimental populations is the most evolutionarily relevant quantity, representing the fraction of mutations that can occur and become established before the back mutation.

Sequencing intragenic mutations: From each replicate population that recovered fitness by acquiring a compensatory mutation, we sequenced an interval of the genome enclosing the gene that contains the deleterious mutation to identify intragenic compensatory mutations. The types and locations of nucleotide substitutions from this sequencing are summarized in Table 2. We identified a total of 55 unique intragenic compensatory mutations that occurred at least once in our sample of 172 genotypes. No more than one intragenic mutation was observed in any fitness-recovering genotype. An absence of any substitution in the sequenced interval was interpreted as representing an instance of extragenic compensatory mutation. In total, there are 78 putative cases of extragenic compensatory mutation (Table 2), which compose an average proportion of 46% for a random sample of compensatory mutants overall. This implies that epistatic interactions occur between mutations in different genes almost as often as within genes in this organism. All of the intragenic mutations observed cause a missense amino acid substitution in the gene containing the deleterious mutation, with the exception of T121C and C144T (Table 2), which are silent sub-

TABLE 2

Summary of intragenic sequencing results

Mutation	Intragenic substitutions	Amino acid substitution	No. of putative extragenic substitutions
G232C	C144T(3), G265A(2), *T367C*(6)	F4F, V45I, Stop79Q	3
G1132C	A1310G(1), G1319A(1), *G1428T*(1), *T2085C*(1), G2179C(1)	I104V, A107T, G143V, V362A, Q393H	7
C1137T	G2050T(1)	Q350H	12
G1423C	*G1428A*(7), A2146T(1), G2150A(1)	G143D, E382D, V384I	1
A1624T	A1250G(1), G1252T(2), G1274A(1), *G1428A*(4), T1606A(1), A1623C(1), G1972T(1), T2030C(1)	K84E, K84N, A92T, G143D, N202K, E208A, L324F, F344L	1
T1631C	—	—	2
T1757A	G1178A(2), A1883G(1), *T2085C*(1)	A60T, T295A, V362A	1
C2235T	—	—	5
C2480T	—	—	15
G2746C	A2551G(3), A2652T(1), A2753G(2), C2740T(1), A2854G(1)	N53D, K86N, Y120C, H115Y, K154E	0
G2979A	G3168T(1)	G80C	2
A3054C	C3016T(1), A3069G(1), A3100G(1), G3111A(3), *G3121T*(4), A3124G(1), G3132T(1)	A29V, M47V, Q57R, V61I, R64L, D65G, A68S	4
T3288C	G3484A(1), C3500G(1), C3511T(1), G3524T(1)	R185H, D190E, A194V, M198I	0
G3326T	C3190T(1), A3274G(1), C3331T(1), G3336A(1), C3337T(7)	T87M, K115R, A134V, A136T, A136V	3
G3524C	A3541G(1), A3577G(1), C3562T(1), G3636T(1), A3696G(2), G3708A(1), A3709G(1), A3819G(1)	K204R, N216S, A211V, A236S, S256G, D260N, D260G, I297V	7
A5031G	A4899G(1), G5042A(4)	I307V, M354I	3
C5337A	G55T(1), T121C(1)	E487D, T509T	12

Substitutions in italics indicate reversions of mutations that had accumulated during adaptation of *300aI* (indicated by underlining) or divergence of the wild-type ancestor from the Sanger *et al.* (1978) genotype. The number of independent isolates for each substitution is given in parentheses.

stitutions in the corresponding genes that contain the deleterious mutation (*i.e.*, genes *A* and *C*, respectively). However, both T121C and C144T occur in a regions of genes *A* and *C* that overlap the gene *K*, causing missense and nonsense substitutions in the latter (C24R and R32Stop, respectively), and therefore are not truly silent substitutions. Indeed, they may represent cases of extragenic compensatory mutation. No function has yet been attributed to gene *K*, although it is nonessential for growth (Hayashi *et al.* 1988).

Several of the intragenic compensatory mutations involve sites in the genome implicated in the adaptation of 300*aI* in the laboratory (G1428A, G1428T, and G3121T) or in the divergence of the wild-type genotype from the Sanger *et al.* (1978) sequence (T367C, T2085C, and G3121T). The reversion of substitutions implicated in the adaptation of ϕX174 in the laboratory as subsequent compensatory mutations is consistent with our observation that compensatory mutations have deleterious effects when isolated in the ancestral background (see below). Two of these compensatory mutations (G1428A and T2085C) were observed in replicate populations derived from more than one deleterious mutation. In other words, either of these mutations may be capable of acting globally on a wide variety of deleterious mutations occurring in gene *F*, which encodes the major capsid protein of ϕX174. Both of these compensatory mutations reside on the structures that compose the exterior of the viral capsid (affecting residues 143 and 362 of the F protein on insertions between β-strands E and F and H and I, respectively; McKenna *et al.* 1992). Most points of contact between F proteins in the formation of the pentameric structural component of the viral capsid are located in these external structures, and likewise most temperature-sensitive mutations that impede procapsid assembly are found here (McKenna *et al.* 1992). In fact, all but one of the compensatory mutations affecting the deleterious mutation A1624T (Table 2), including G1428A, are found on these external structures.

Severe mutations are more compensable: There is an abundance of variation in the relative fitness of the deleterious mutations in our study, ranging from 0.04 to 79% of the average yield of ancestral 300*aI* phage after one growth cycle (Table 1). This presents a unique opportunity to evaluate the relationship between the severity of a deleterious mutation and its associated rate of compensatory mutation. Figure 2 displays the observed proportion of replicate populations that recover from the effects of a deleterious mutation by acquiring

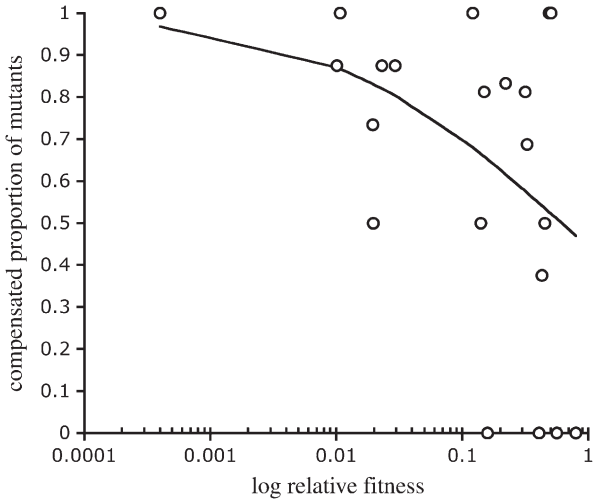

FIGURE 2.—Plot of the observed proportion of replicate populations that independently acquire a compensatory mutation, instead of the back mutation, as a function of the log-transformed relative fitness of the deleterious mutation. Log relative fitness is a significant factor ($P < 10^{-4}$) in a binomial regression model, which predicts that the proportion of replicates with compensatory mutations increases as the deleterious mutation becomes more severe. The solid line indicates the expected proportion predicted by the binomial regression model as a function of log relative fitness. A regression analysis carried out for the estimated number of intragenic compensatory mutations, rather than the proportion of replicates that acquire any compensatory mutation, obtains a similar result (discussed in text).

a compensatory instead of back mutation, plotted against the log-transformed relative fitness of each deleterious mutation. We applied a binomial regression in which the proportion of replicates that acquire a compensatory mutation is, on a logit scale, a linear function of log relative fitness. This regression specifies a highly significant negative effect of log relative fitness ($P < 10^{-5}$) on the proportion of compensated replicates, implying that populations afflicted by more severe deleterious mutations are more likely to recover by acquiring compensatory mutations. The expected probability of compensatory mutation as a function of mutation severity, as predicted by the regression model, is indicated by the line in Figure 2. The binomial regression model explains a significant proportion of variation in the observed proportion of replicates with compensatory mutations ($\chi^2 = 25.23$, d.f. $= 1$, $P < 10^{-5}$). However, a substantial amount of variance remains unexplained by the model (residual $\chi^2 = 135.87$, d.f. $= 19$, $P < 10^{-5}$). Uncertainty in the estimate of the compensated proportion from a limited number of trials is undoubtedly a major source of variance.

The above analysis examines the effect of the severity of the deleterious mutation on the proportion of replicate populations acquiring any compensatory mutation. Although the latter quantity corresponds to how compensatory mutations are revealed to evolution, it cannot address directly the mechanistic relationship between mutation severity and the number of compensatory mutations. However, in place of the proportion of replicates with a compensatory mutation, we can instead estimate the total number of different intragenic compensatory mutations on the basis of the number of mutations observed in sequencing (Table 2). Although the latter quantity is potentially more revealing, it is also more difficult to apply. For instance, when every replicate in a sample has independently acquired a unique compensatory mutation, as in the case of T3288C (Table 2), it is nearly impossible to arrive at a meaningful estimate of the actual number of compensatory mutations. As a result, it was necessary to remove cases in which our sample of intragenic compensatory mutations is far from saturation (*e.g.*, C1137T, G2979A, T3288C, G3524C, and C5337A). For the remaining cases, we employed a probability formula assuming random sampling (READ 1998, Equation 4) to obtain a maximum-likelihood estimate of the total number of compensatory mutations. In a Poisson logistic regression, we detected a significantly negative relationship between the log relative fitness of a deleterious mutation and the maximum-likelihood estimate of the total number of intragenic compensatory mutations ($P < 0.005$). This implies that the previous trend for increasing compensability with mutation severity extends to the number of compensatory interactions within the gene. Overall, the relationship between mutation severity and compensability has intriguing implications for the biological basis of epistatic interactions. We address these implications and issue some caveats in measuring this relationship in the DISCUSSION.

Clustering of intragenic mutations: The intragenic compensatory mutations identified by sequencing are often found in close proximity to the deleterious mutation in the linear amino acid sequence (Table 2). We quantified the degree of clustering in the linear sequence by calculating the absolute codon distance that separates each pair of compensatory and deleterious mutations, normalized by the gene length. For our data, this test statistic has an average value of $\sim$0.191. In other words, the average compensatory mutation is located no farther than 20% of the gene length from the deleterious mutation in the linear sequence. To evaluate whether this average distance was shorter than expected from a random distribution of mutations, we randomized the locations of compensatory mutations to generate a large number ($N = 10^5$) of random outcomes. The resulting null distribution of the test statistic has an average value of 0.303 (95% C.I.: 0.259–0.347). Hence, compensatory mutations are located significantly closer to the deleterious mutation than expected from a random distribution.

We also located the residues affected by compensatory mutations in the published protein structure of the ϕX174 procapsid (DOKLAND *et al.* 1997). To facilitate interpretation, we will indicate the amino acid residue

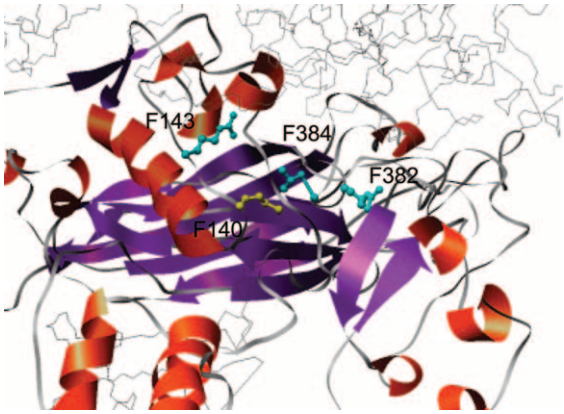

Figure 3.—Compensatory mutations of G1423C mapped to residues in the F protein structure. The residue affected by the deleterious mutation G1423C (*i.e.*, F140) is highlighted in yellow, and compensatory residues are highlighted in blue. Each pair of residues is no more than 15 Å apart. These residues form a group on the external face of the F protein, located above the β-barrel indicated by the purple ribbon. This image was generated in UCSF Chimera v. 1.2 (Computer Graphics Laboratory, UCSF) using the published structure PDB ID 1AL0 (see MATERIALS AND METHODS).

affected by the deleterious or compensatory mutation by the gene letter followed by residue number (*e.g.*, F140 refers to residue 140 in gene *F*). In the folded structure of the procapsid, many compensatory mutations are located closer to the deleterious mutation than implied by the pairwise distance in the linear sequence. Figure 3 illustrates a particular instance of this structural clustering in which compensatory mutations (at positions F143, F382, and F384) that interact with the deleterious mutation G1423C (F140) map to residues that form a tight cluster in the folded F protein structure. In this case, the clustering test statistic for G1423C is misleadingly high (0.382). Of course, not all compensatory mutations cluster around the deleterious mutation in protein structures. For example, the deleterious mutation A1624T (F207), which maps to an α-helix in the EF insertion (as described above), is affected by five compensatory mutations that modify residues located on or near this α-helix (*e.g.*, at positions F143, F202, F208, F324, and F344) and three other mutations on the far end of the insertion (*e.g.*, at positions F84 and F92). Because this structure is implicated in establishing contacts with neighboring F proteins in the pentameric unit, it is possible that the latter group of mutations represent compensatory interactions between subunits.

Analysis of intragenic mutation diversity: We applied a likelihood analysis to our sequencing results to obtain an estimate for the average number of different intragenic compensatory mutations that affect a given deleterious mutation (herein referred to as $E[C_{\text{intra}}]$). The number of intragenic compensatory mutations that occur at least once in a random sample provides information about the total number of mutations. For example, the multiple occurrence of three different intragenic mutations (C144T, G265A, and T367C) in 11 replicate populations fixed for the deleterious mutation G232C (Table 2) implies that sampling of mutations is near saturation. In other words, we infer that additional novel mutations are unlikely to appear with further sampling. On the other hand, the four compensated replicate populations fixed for T3288C had each acquired a unique intragenic compensatory mutation, implying that other mutations are possible. A likelihood model based on this property of random sampling can determine what hypothetical distribution of C_{intra} best explains the data. This likelihood analysis was developed previously for a study on estimating the overall statistical distribution of C for a wide range of genes and taxa (Poon *et al.* 2005, this issue). Here, our maximum-likelihood estimate for the distribution of C_{intra} is described by an L-shaped gamma distribution function ($\alpha = 0.383$, $\theta = 23.41$) that predicts an average of $E[C_{\text{intra}}] = 8.96$ intragenic mutations. This analysis assumes uniform mutation rates among sites within a gene and will underestimate C_{intra} when this assumption is violated.

Using a randomization analysis, we determined that this distributional estimate predicts that the replicate populations in our experiment would acquire intragenic compensatory mutations instead of back mutations with an average frequency of 54.4% (95% C.I.: 40.6 and 68.2%). We expect improved concordance with our observed frequency of 67.7% if we incorporate extragenic compensatory mutation and variation in mutation rates. The shape of the distribution of C plays an important role in relating the average, $E[C]$, to the frequency of compensatory mutation in replicate populations. A bell-shaped distribution with low variation in the number of compensatory mutations would constrain the average to only two or three, to be consistent with an observed frequency of ~70%. In contrast, an L-shaped distribution stipulates greater variation in C and thereby predicts a greater average $E[C]$ for the same observed frequency of compensatory mutations.

Compensatory mutations are deleterious: When we classify fitness-recovering mutations that occur at sites apart from the deleterious mutation as compensatory, we implicitly assume that there is a strong epistatic interaction that is principally responsible for the recovery of fitness. However, an alternate possibility is that fitness is being recovered by mutations that are beneficial irrespective of their genetic context, which can be indistinguishable from compensatory mutations. As an initial precautionary measure, we evolved the ancestral population by serial transfer for 300 growth cycles under controlled conditions. However, the evolved 300*aI* genotype does not necessarily occupy a fitness optimum that precludes unconditionally beneficial mutations.

To directly address this possibility, we used site-directed mutagenesis to revert the deleterious mutation in an arbitrary sample of evolved genotypes that were

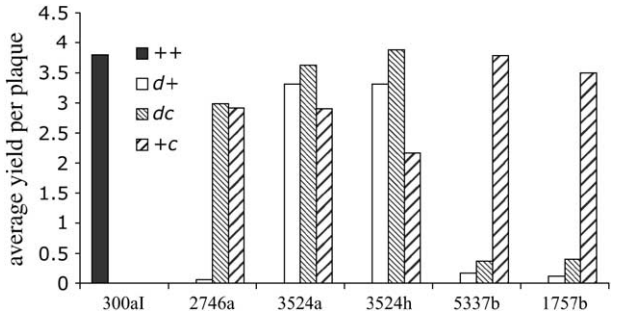

FIGURE 4.—Fitness estimates for sets of deleterious and compensatory mutant genotypes, with respect to average yield per plaque. Each set of genotypes for a pair of deleterious and compensatory mutations is denoted by the location of the deleterious mutation and replicate population designation letter. Specifically, 3524*a* refers to the deleterious mutation G3524C and the intragenic compensatory mutation G3562A, which appeared in replicate population *a*. The other compensatory mutations are as follows: 3524*h*, T3696C; 2746*a*, A2652T; 1757*b*, A2404G (an extragenic mutation in gene *G* isolated in two independent replicates); and 5337*b*, A121G. Fitness estimates are averaged over six or more replicate assays for each genotype. The symbol "+" denotes the wild-type allele that occupies the site of either the deleterious (*d*) or the compensatory (*c*) mutation. Fitness of the ancestor 300*aI* (++) is represented by a solid bar. Open bars represent deleterious mutations (*d*+), thinly hatched columns are compensated deleterious genotypes (*dc*), and thickly hatched columns are isolated compensatory mutations (+*c*).

identified as containing both deleterious and compensatory mutations. In other words, we moved the putative compensatory mutations into the ancestral (300*aI*) genetic background. Subsequently, we measured the average yield per plaque for all four combinations of two biallelic loci: namely, the ancestor 300*aI* (++), the deleterious mutation with (*dc*) and without (*d*+) the compensatory mutation, and the compensatory mutation itself (+*c*). The results of this assay are presented in Figure 4. In all cases, the compensatory mutation has a deleterious or negligible effect on fitness when placed in the ancestral background. This is consistent with the recovery of fitness being caused by an epistatic interaction with the deleterious mutation, rather than by an unconditionally beneficial effect of the second mutation itself. To address the possibility that this result was due to the spontaneous occurrence of mutations during reversion of deleterious mutations by site-directed mutagenesis, fitness estimates were assayed for multiple phage isolates from the transformation mixture. No significant difference in fitness estimates was observed among replicate transformation isolates (data not shown).

DISCUSSION

Adaptation and compensatory mutations: We estimate that, on average, ~70% of replicate populations recovering from the effects of a fixed deleterious mutation will acquire a compensatory mutation instead of a back mutation in the bacteriophage ϕX174. Moreover, our analysis of the diversity of intragenic compensatory mutations, unearthed by sequencing evolved genotypes from the replicate populations, predicts that any deleterious mutation can be compensated by at least nine other mutations on average. We note that there is substantial variance around this mean value (described by an L-shaped gamma distribution); indeed, some deleterious mutations in our study evidently cannot be compensated for. Our results make a strong overall case that compensatory mutation plays an important, and as yet neglected, role in biological evolution. For example, LEVIN *et al.* (2000) have demonstrated that bottlenecked populations tend to fix compensatory mutations, instead of back mutations, when the rate of compensatory mutation exceeds the back mutation rate by at least an order of magnitude. There is an important stipulation, however, to formulating a direct analogy from our experimental populations to biological reality: deleterious mutations as severe as those used in our experiment are unlikely to become fixed under most natural conditions. In spite of this, our experiment reveals a remarkable abundance and diversity of epistatic interactions that not only engender compensatory mutations under artificial conditions, but also represent multiple alternative solutions to adaptation under natural conditions. In other words, a deleterious mutation is genetically equivalent to the absence of a beneficial mutation. A natural population will therefore be placed in a situation analogous to our experiment following a change in the environment that necessitates adaptation. For instance, pathogens can rapidly acquire a mutation that confers resistance to an inhibitory drug, but the mutation often becomes deleterious when the population is returned to the drug-free environment (MAISNIER-PATIN and ANDERSSON 2004).

Our results have several implications for the genetic complexity of adaptation. First, we predict that populations recently derived from a common ancestor will tend to respond to selection in the same novel environment by acquiring divergent sets of mutations. A corollary of this prediction is that when a population is returned to the ancestral environment, it is less likely to return to the ancestral genotype. Second, we predict that cases of adaptation that involve mutations with large effects are more likely to promote genetic divergence among populations. Finally, we predict that genetic divergence will involve mutations in different genes almost as often as in the same gene. Presently, few experimental studies can unambiguously verify or contradict these predictions (PORTER and CRANDALL 2003). One notable exception is a study by CRILL *et al.* (2000) in which populations of ϕX174 were subjected to alternating host environments (*E. coli* and *Salmonella enterica*). One site in the F protein repeatedly acquired reversions in a correlated response to changing hosts and was later

determined to have a major effect on the rate of host-specific growth. However, only three of seven mutations acquired in the initial adaptation of the ancestral population to *S. enterica* were reverted in both replicate populations that were subsequently derived from the ancestor (Crill *et al.* 2000). Unfortunately, it is unclear whether the absence of correlated reversions in this study is caused by compensatory mutations. Further experiments will be required to elucidate how often epistasis drives the divergence of populations during adaptation.

Mutation severity and compensability: We have determined that the frequency at which populations of ϕX174 acquire compensatory mutations instead of back mutations to recover fitness is a significantly increasing function of the severity of the deleterious mutation. This trend evidently extends to the total number of intragenic compensatory mutations, which also increases with deleterious mutation severity. It is possible, however, that the observed relationship between mutation severity and frequency of compensatory mutation is an artifact of our experimental design. First of all, mutations that recover fitness in a population fixed for a weakly deleterious mutation are difficult to detect. Because it may be necessary to evolve such populations for a greater number of generations, there may be greater opportunity for a back mutation to occur and out-compete the compensatory mutation. Second, if the effect size distribution of compensatory mutation scales with the severity of the deleterious mutation, then for weaker deleterious mutations a greater proportion of compensatory mutations could drop below the detection threshold.

We have been unable to recover either back or compensatory mutations for the weakly deleterious mutations A3582G and G4970T. However, we have been able to isolate a small number of back and compensatory mutations from two other deleterious mutations, C1280T and T1631C, with comparably weak effects on fitness. Because these were isolated after <25 growth cycles, there was insufficient time for a back mutation to arise and interfere with the spread of a compensatory mutation (Gerrish and Lenski 1998). Note that we did not require mutations to become fixed in the population. Furthermore, even if all fitness-recovering mutations isolated for C1280T or T1631C were caused by compensatory mutations, the regression would remain highly significant. On the other hand, the deleterious mutations T489C and T495C have a strong influence on our regression. We have recovered only back mutations for these two mutations, which have weak effects on fitness with respect to the average yield per plaque. However, both T489C and T495C also create plaques that are disproportionately small (Table 1), so that the range of plaque size above the detection threshold is comparable to most other mutations in our study. Moreover, no more than 10 and 22 growth cycles were required to detect fitness-recovering mutations from T489C and T495C, respectively. Therefore, we infer that for the range of fitness effects in our study, the detection threshold is not the principal cause of the observed severity-compensability relationship. Also, the compensatory mutations that fall below the detection threshold are the least likely to contribute to the evolution of fitness because they cannot compete against the back mutation, as noted above.

Notwithstanding experimental biases, what could be the biological cause by which more severe deleterious mutations tend to create a greater number of potential sites for compensatory mutation? One intriguing possibility is that deleterious mutations with large effects on fitness may tend to affect a broader range of phenotypic components. Severely deleterious mutations would therefore generate a larger mutational target for compensatory interactions. This notion conceivably applies at any biological level of organization. For instance, a mutation that causes a stronger perturbation in the overall structure of a folded protein may interact with a larger number of residues. Similarly, a mutation at a node of a gene network with higher connectivity by definition interacts with subsequent mutations at a larger number of nodes. An increase in the number of compensatory mutations with increasing distance from a fitness optimum is also predicted by Fisher's (1930) geometric model of adaptation. When a population is close to a fitness optimum, its fitness can be improved only by a narrow range of mutations of small effect. However, if the population is far from an optimum, then it may indiscriminately select from a wide range of mutations, even from those associated with deleterious pleiotropic side effects. This relationship between mutation severity and compensability has an unfortunate implication, however, in that the deterioration of fitness from the accumulation of deleterious mutations is largely caused by mutations of slight effect (Kondrashov 1995). Consequently, compensatory mutation may be unable to prevent the mutational meltdown of all but the smallest populations. On the other hand, compensatory mutation represents an extreme form of antagonistic epistasis. If an overall decline in antagonistic epistasis with the effect size of deleterious mutations is accompanied by an increase in synergistic epistasis, then a lack of compensatory mutations may be offset by an increased efficiency of selection for removing slightly deleterious mutations (Kondrashov 1994).

The intragenic-extragenic ratio: A substantial proportion (~46%) of our replicate populations have evidently acquired extragenic compensatory mutations to recover fitness. It is also conceivable that some of the intragenic mutations that we have identified by sequencing are in fact neutral in effect, causing us to underestimate the proportion of compensatory mutations that are extragenic. Because we isolate fitness-recovered plaques as soon as they are detected, however, there is very limited opportunity for a neutral mutation to rise to a

frequency sufficiently high for this to occur. Furthermore, the significant degree of clustering and multiple independent occurrences implies that at least most of the observed intragenic mutations are indeed compensatory. The high frequency of extragenic compensatory mutation that we have observed is corroborated by several studies of suppressor mutations in the bacteriophage ϕX174. These suppressor mutations reveal functional interactions within and among gene products, particularly the proteins that participate in the assembly of the viral procapsid (*i.e.*, genes *B*, *D*, *F*, and *G*; Dokland *et al.* 1997). For example, although the internal scaffolding gene *B* is essential, deletion of nearly half of *B* can yield viable, albeit defective, bacteriophage, demonstrating functional redundancy among genes (Novak and Fane 2004). Indeed, the remaining defective phenotype in deletion mutants can be ameliorated by extragenic suppressor mutations in gene *D*. Furthermore, *B* and *D* may represent an evolutionary partition of functions carried out by a single scaffolding protein in other systems (Fane and Prevelige 2003). Extragenic suppressor mutations are also common among other gene products in ϕX174. Mutations that disrupt the DNA-binding function of the *J* gene product are suppressed by mutations in the major capsid gene *F* (Jennings and Fane 1997; Hafenstein *et al.* 2004). Similarly, suppressors of a cold-sensitive mutation in gene *D* were located in either gene *F* or gene *J* (Fane *et al.* 1993). Hence, there is a multitude of associations among different gene products that participate in the procapsid.

Extragenic suppressor mutations are also frequently observed in organisms other than bacteriophage. A survey of the suppressor mutation literature estimates that the overall proportion of mutations that are extragenic is ~10% for all taxa (Poon *et al.* 2005, this issue). This proportion, however, increases to ~35% if calculated specifically for mutations in viruses; hence, our result is consistent with observations in other viral systems. Why are extragenic mutations evidently more common in viruses than in prokaryotes or eukaryotes? We would expect the opposite trend, because prokaryotic and eukaryotic genomes generally consist of a considerably greater number of genes, potentially conferring a larger mutational target for extragenic interactions. On one hand, extragenic mutations may be underreported in studies of prokaryotes and eukaryotes. However, such studies are principally motivated to identify and characterize functional interactions among different proteins (Sujatha and Chatterji 2000). On the other hand, this trend could reflect a lesser degree of compartmentalization of biological function among genes in viruses than in prokaryotes or eukaryotes. Viruses are under strong selection for genomic compactness; for instance, several genes in the bacteriophage ϕX174 overlap or are nested within one another (Hayashi *et al.* 1988). It is plausible that such adaptations in genome organization restrict the ability for viral genes to functionally specialize, which could limit the number of extragenic interactions in higher organisms.

Future directions: Our understanding of adaptation and the evolutionary consequences of mutation will benefit greatly from further empirical investigations into the phenomenon of compensatory mutation. Can populations that have acquired a compensatory mutation in lieu of reverting a deleterious mutation recover fitness completely, or are they trapped on an inferior "fitness peak"? What is the effect size distribution of compensatory mutations, and does this differ between mutations located intragenically and extragenically? Does the accumulation of deleterious mutations enhance or impede the recruitment of potential sites for compensatory mutations in the genome? The bacteriophage ϕX174 provides an ideal experimental system in which to address these and other problems in evolutionary biology. Both this study and previous work with suppressor mutations (*e.g.*, Jennings and Fane 1997) demonstrate the versatility of site-directed mutagenesis in this system. However, because ϕX174 exhibits rudimentary gene regulation (Hayashi *et al.* 1988), studies of epistatic interactions in this bacteriophage are limited to those within or among structural gene products. To address the nature of compensatory mutations in a regulatory gene network will require the development of a similar experimental system in a more biologically complex model organism.

A promising extension of our investigations in compensatory mutation is found in recent theoretical work on the effect size distribution of beneficial mutations available for adaptation (Orr 2003). By assuming that genotypes in natural populations tend to occupy the upper limit of a fitness distribution, Orr (2003) could apply extreme value theory to demonstrate that the difference in fitness separating adjacent genotypes ranked by fitness is exponentially distributed. This provides a testable hypothesis that can be evaluated for compensatory mutations in bacteriophage ϕX174 (Rokyta *et al.* 2005). To provide a deeper understanding of the adaptive landscape, Orr's (2003) result might also be combined with our observation that the number of compensatory mutations available for evolution is a decreasing function of the distance from the fitness optimum.

We thank B. A. Fane for generously providing the wild-type ϕX174 and *E. coli C* host strains and protocols; H. Hoekstra and R. Burton for providing sequencing resources; Lawrence K. Chao for laboratory assistance; and B. H. Davis, H. Hoekstra, D. M. Weinreich, J. J. Bull, and an anonymous reviewer for their insightful comments on previous versions of this article. This work was supported by a grant from the National Institutes of Health (GM-60916) to L.C.

LITERATURE CITED

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat *et al.*, 2000 The Protein Data Bank. Nucleic Acids Res. **28:** 235–242.

Burch, A. D., and B. A. Fane, 2000 Foreign and chimeric scaffolding proteins as inhibitors of *Microviridae* morphogenesis. J. Virol. **74:** 9347–9352.

Burch, C. L., and L. Chao, 1999 Evolution by small steps and rugged landscapes in the RNA virus ϕ6. Genetics **151:** 921–927.

Burch, C. L., and L. Chao, 2004 Epistasis and its relationship to canalization in the RNA virus ϕ6. Genetics **167:** 559–567.

Chao, L., T. T. Tran and T. T. Tran, 1997 The advantage of sex in the RNA virus ϕ6. Genetics **147:** 953–959.

Crill, W. D., H. A. Wichman and J. J. Bull, 2000 Evolutionary reversals during viral adaptation to alternating hosts. Genetics **154:** 27–37.

Dokland, T., R. McKenna, L. L. Ilag, B. R. Bowman, N. L. Incardona *et al.*, 1997 Structure of a viral procapsid with molecular scaffolding. Nature **389:** 308–313.

Fane, B. A., and M. Hayashi, 1991 Second-site suppressors of a cold-sensitive prohead accessory protein of bacteriophage ϕX174. Genetics **128:** 663–671.

Fane, B. A., and P. E. Prevelige, Jr., 2003 Mechanism of scaffolding-assisted viral assembly. Adv. Protein Chem. **64:** 259–299.

Fane, B. A., S. Shien and M. Hayashi, 1993 Second-site suppressors of a cold- sensitive external scaffolding protein of bacteriophage ϕX174. Genetics **134:** 1003–1011.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection*, Ed 2. Dover Publications, New York.

Gerrish, P. J., and R. E. Lenski, 1998 The fate of competing beneficial mutations in an asexual population. Genetica **102/103:** 127–144.

Hafenstein, S. L., M. Chen and B. A. Fane, 2004 Genetic and functional analysis of the ϕX174 DNA binding protein: the effects of substitutions for amino acid residues that spatially organize the two DNA binding domains. Virology **318:** 204–213.

Hayashi, M., A. Aoyama, D. L. Richardson, Jr. and M. N. Hayashi, 1988 Biology of the bacteriophage ϕX174, pp. 1–71 in *The Bacteriophages*, edited by R. Calendar. Plenum Press, New York.

Jennings, B., and B. A. Fane, 1997 Genetic analysis of the ϕX174 DNA binding protein. Virology **227:** 370–377.

Kimura, M., 1985 The role of compensatory neutral mutations in molecular evolution. J. Genet. **64:** 7–19.

Kondrashov, A. S., 1994 Muller's ratchet under epistatic selection. Genetics **136:** 1469–1473.

Kondrashov, A. S., 1995 Contamination of the genome by very slightly deleterious mutations: Why have we not died 100 times over? J. Theor. Biol. **175:** 583–594.

Levin, B. R., V. Perrot and N. Walker, 2000 Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics **154:** 985–997.

Maisnier-Patin, S., and D. I. Andersson, 2004 Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. Res. Microbiol. **155:** 360–369.

McKenna, R., D. Xia, P. Willingmann, L. L. Ilag, S. Krishnaswamy *et al.*, 1992 Atomic structure of single-stranded DNA bacteriophage ϕX174 and its functional implications. Nature **355:** 137–143.

Novak, C. R., and B. A. Fane, 2004 The functions of the N terminus of the ϕX174 internal scaffolding protein, a protein encoded in an overlapping reading frame in a two scaffolding protein system. J. Mol. Biol. **335:** 383–390.

Orr, H. A., 2003 The distribution of fitness effects among beneficial mutations. Genetics **163:** 1519–1526.

Orr, H. A., and J. A. Coyne, 1992 The genetics of adaptation: a reassessment. Am. Nat. **140:** 725–742.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt *et al.*, 2004 UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. **25:** 1605–1612.

Poon, A., B. H. Davis and L. Chao, 2005 The coupon collector and the suppressor mutation: estimating the number of compensatory mutations by maximum likelihood. Genetics **170:** 1323–1332.

Porter, M. L., and K. A. Crandall, 2003 Lost along the way: the significance of evolution in reverse. Trends Ecol. Evol. **18:** 541–547.

Read, K. L. Q., 1998 A lognormal approximation for the collector's problem. Am. Stat. **52:** 175–180.

Rokyta, D. R., P. Joyce, S. B. Caudle and H. A. Wichman, 2005 An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nat. Genet. **37:** 441–444.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., A. R. Coulson, T. Friedmann, G. M. Air, B. G. Barrell *et al.*, 1978 The nucleotide sequence of bacteriophage ϕX174. J. Mol. Biol. **125:** 225–246.

Sujatha, S., and D. Chatterji, 2000 Understanding protein-protein interactions by genetic suppression. J. Genet. **79:** 125–129.

Whitlock, M. C., and S. P. Otto, 1999 The panda and the phage: compensatory mutations and the persistence of small populations. Trends Ecol. Evol. **14:** 295–296.

Communicating editor: D. Begun