Experimental Genomic Evolution: Extensive Compensation for Loss of DNA Ligase Activity in a Virus

D. Rokyta, M. R. Badgett,* I. J. Molineux,†‡ and J. J. Bull*‡*

*Section of Integrative Biology, †Section of Molecular Genetics and Microbiology, and ‡Institute of Cellular and Molecular Biology, University of Texas, Austin

Deletion of the viral ligase gene drastically reduced the fitness of bacteriophage T7 on a ligase-deficient host. Viral evolution recovered much of this fitness during long-term passage, but the final fitness remained below that of the intact virus. Compensatory changes occurred chiefly in genes involved in DNA metabolism: the viral endonuclease, helicase, and DNA polymerase. Two other compensatory changes of unknown function also occurred. Using a method to distinguish compensatory mutations from other beneficial mutations, five additional substitutions from the recovery were shown to enhance adaptation to culture conditions and were not compensatory for the deletion. In contrast to the few previous studies of viral recovery from deletions, the compensatory changes in T7 did not restore the deletion or duplicate major regions of the genome. The ability of this deleted genome to recover much of the lost fitness via mutations in its remaining genes reveals a considerable evolutionary potential to modify the interactions of its elements in maintaining an essential set of functions.

Introduction

A genome is a collection of genes that directs and coordinates the functions essential to an organism through a dynamic interaction of its elements. Advances in molecular biology have made it possible to begin understanding how different components of a genome function together. Microarrays, two-hybrid systems, engineered genome reorganizations, and even the simple analysis of loss-of-function mutations are revealing the nature of gene–gene interactions, the contributions of different genes to phenotypes, and redundancies among the genes (Chien et al. 1991; Goebl and Petes 1996; Brown and Botstein 1999; Wagner 2000; Jeong et al. 2001; Peterson and Fraser 2001; Raamsdonk et al. 2001). Biology may thus be on the brink of understanding how simple genomes work.

An even more challenging problem is to understand how a genome evolves. Genome evolution depends not only on how the elements of a genome function and interact with each other but also on how mutations can change those functions and interactions. We have no easy assay of function that allows us to assess the overall role of different genes within the context of a complete genome, so the identification and significance of genetic changes during evolution has been laborious. Furthermore, although genomes evolve by a combination of point mutations and wholesale gene additions and losses or rearrangements, the contributions of these different scales of mutations to genome function are not easily delineated.

Experimental evolution offers a direct means of studying mechanisms of genome interaction and evolution. A simple, yet powerful approach of this sort is the selection of second-site suppressors that compensate for a debilitating mutation. A second-site suppressor provides information about potential genome evolution because it identifies an evolutionary pathway that compensates for or restores a missing function. Genome evolution encompasses a wide range of mechanisms and multiple events that are not likely to arise through natural mechanisms in the course of short-term experiments, such as gene duplications, rearrangements, and horizontal gene transfer. Yet the ability to engineer profound genome manipulations in many organisms now allows one to study experimentally the evolutionary consequences of many types of major genomic mutations that were not possible previously (Ball et al. 1999; Berkhout et al. 1999; Endy et al. 2000; Jackson et al. 2001).

Within this framework, experimental studies of viruses containing major deletions or rearrangements offer some of the first studies of genome evolution in response to major genomic changes (Whatmore et al. 1995; Olsthoorn and van Duin 1996; Berkhout et al. 1999). These studies used evolution not only to understand genome organization but also to help assemble and improve artificially created genomes. Here we examine the genomic response to a deletion of the DNA ligase gene of bacteriophage T7. Although the phage ligase is not essential for T7 growth in the presence of normal host ligase activity, when the host ligase is defective, a deletion of phage gene is effectively lethal. Subsequent passage of the crippled T7 was undertaken to determine if and how the virus would recover fitness through compensatory evolution of its 58 other genes.

Materials and Methods

Phage, Bacteria, and Plasmids

The virus used in this study was wild-type bacteriophage T7, with a 40-kbp linear genome of double-stranded DNA. Standard manipulations of T7 were those described by Studier (1969). The bacterial host used for evolution was strain IJ434, a derivative of *Escherichia coli* B that carries a defective ligase (Studier 1973); IJ434 also carries a mutation (*optA1*) enhancing levels of dGTPase (Beauchamp and Richardson 1988), but the

Key words: microbe, evolution, bacteriophage, genomics, compensatory evolution, resistance.

Address for correspondence and reprints: J. J. Bull, Section of Integrative Biology, University of Texas, Austin, Texas 78712. E-mail: bull@bull.biosci.utexas.edu.

Mol. Biol. Evol. 19(3):230–238. 2002
 ISSN: 0737-4038

optA1 mutation is not thought to be important for this study. Normal T7 growth using IJ434 as host requires the activities of genes *1.2* (anti-dGTPase) and *1.3* (DNA ligase). The *E. coli* K-12 Su^+ strain IJ511 (*E. coli* K-12 ΔlacX74 supE44 galK2 galT22 mcrA rfbD1 mcrB1 hsdS3) was used to propagate T7 amber mutants, IJ1126 (*E. coli* K-12 *recB21 recC22 sbcA5 endA gal thi* Su^+ Δ(*mcrC-mrr*)*102::*Tn*10*) was used for transfection of genomic T7 DNA. The plasmid pAR1798, kindly provided by Dr. F. W. Studier, Brookhaven National Laboratories, carries the T7 gene *1.3* flanked by its promoter *ø1.3* and the early transcription terminator TE (T7 coordinates 15.93–19.31) inserted at the *Bam*HI site of pBR322.

Site-directed mutagenesis was used to introduce individual mutations into the T7 genome. This procedure, which was used to introduce both point mutations and deletions, involved creating a plasmid in which the desired mutation was flanked on each end by 50–100 bp T7 DNA from the genomic location where the mutation was to be inserted. T7 was plated on hosts carrying the plasmid, phages within a plaque were then replated directly on a selective host or on a plasmid-free host for screening. Plasmids with deletions were created either by (1) self-ligating inverse PCR products of plasmids containing the full T7 gene sequence (the deleted DNA being excluded during the PCR), or by (2) amplifying T7 genomic DNA at each end of the deletion as separate products with sequence overlap between 5′ ends of one primer in each reaction, followed by a second PCR to combine the two products. In the latter case, products were cloned directly into Promega pGEM Tvector System I plasmids.

Genomic fragment swaps helped localize mutations that affected phage fitness. DNA from parental and evolved T7 phage was digested with appropriate restriction enzymes, two or three fragments from different genomes were purified, complete sets of fragments were ligated, and reaction products were then transfected into IJ1126. Individual plaques were then tested for phages that could grow on IJ434. Selected regions of phage genomes were then sequenced as PCR products using an ABI 377 and Perkin-Elmer Big-Dye™ reagents. Sequence comparisons were made with DNA Star Lasergene™ (1998).

Media

Liquid media used was LB (10 g NaCl, 10 g Bacto tryptone, 5 g yeast extract per liter), supplemented with $CaCl_2$ to 2 mM. Bottom agar for plates contained 15 g Bacto agar per liter LB broth, and top agar contained 7 g/liter Bacto agar in LB broth.

Phage Propagation

T7 was grown at 37°C in 125 ml flasks containing cells suspended in 10 ml LB broth. Concentrated cell cultures, frozen in aliquots of LB with 20% glyerol, were added to the flasks and grown with aeration for 1–2 h prior to phage addition. Cell density reached approximately $1–2 \times 10^8$ at the time phage were added. Typically, 10^4–10^6 phage were added and grown for 40–60 min. A sample of the culture was then treated with chloroform, and the released phage was used to inoculate the next culture. One passage is thus the amplification of phage between two successive flasks.

The goal in these passages was to select for rapid phage growth while maintaining populations that were large enough to avoid stochastic loss of beneficial mutations. To ensure that the culture conditions favored rapid phage growth, not within-host competition (Turner and Chao 1999), cultures were usually stopped when the phage concentration had reached, at most, 10^9 plaque-forming units per ml, 10-fold or more lower than the concentration usually attained by allowing the culture to proceed to lysis. In the event that phage fitness was too low to ascend four orders of magnitude in 60 min, higher initial concentrations were used or the passage was continued for a longer time.

The fitness of a T7 gene *1.3* deletion mutant on IJ434 is so low immediately after deletion of ligase that it cannot be maintained in flask-to-flask transfers. The phage was therefore grown initially on plates. For the long-term line $T7^{\Delta 1}$, the first passages involved plating enough phage (approximately 10^5–10^7) to produce about 500 moderate-sized plaques over 12–24 h, extracting the phage from top agar over chloroform, and replating. After seven plate passages, the fitness of the phage population was sufficient to allow liquid population. For the short-term lines ($T7^{\Delta 2}$, $T7^{\Delta 3}$), a plaque on IJ434 was picked, repurified by plating on IJ434, and the phage in that plaque used as a stock for fitness assays and genotype determination.

Fitness Measure

T7 fitness was measured simply as the rate of change in phage numbers in a liquid culture of LB under conditions similar to those used for propagation. Fitness assays were conducted in 10 ml LB with IJ434 in 125 ml flasks. The cells were taken from a frozen stock, grown for at least an hour with aeration at 37° to a density of $1–2 \times 10^8$ cells/ml. Typically, 10^2–10^4 phage were added and grown for 40 min (unless fitness was expected to be low, in which case 10^5–10^6 phage were added) before treating the infected cell culture with chloroform. Phage titers were determined both before and after the period of timed growth. Phage stocks were prepared within 2 days of their use in fitness assays. Where necessary, stocks were amplified on *E. coli* B or IJ511 (both of which contain normal ligase activity) instead of IJ434 to avoid selection at this stage for compensatory mutations.

Fitness is presented on a log scale as the growth rate under the standard, defined conditions. If N(0) phage were added to the culture at time zero and N(t) were present at time t (in hours), the ratio N(t)/N(0) represents the factorial increase in phage. For ease in scaling fitnesses, we use $\log_2$ [N(t)/N(0)]/4t, which is the number of doublings of phage numbers per 15 min, the approximate generation time for wild-type T7 on normal hosts. When multiplied by 0.693, this fitness

equals the intrinsic rate of increase used commonly in demographic models, but the term doublings is easier to comprehend for phage growth. Use of doublings per generation does not imply or require that phage reproduction occurs by binary fission, which of course it does not.

In some cases, fitness is expressed as a simple ratio of the titer of the virus on a nonpermissive (restrictive) host to that on a permissive host. This relative efficiency of plating (eop) can serve as a measure of fitness when the eop is $\ll 1$, but it is not useful when the two hosts being considered both propagate the virus and the eop approaches unity.

Lines and Notation

Evolution of the T7 genome in response to deletion of its ligase gene was studied in three independent lines. Prior to the acquisition of the deletion, $T7^{+}$ was preadapted to culture conditions (for 10 and 20 transfers) to help ensure that subsequent adaptation would be in response to the deletion. Deletions of gene *1.3* were introduced into the preadapted virus, and the recombinant virus was then passaged to allow fitness recovery. One line was a long-term adaptation involving many transfers of virus from culture to culture; the other two lines were short term, in which rare plaques on a restrictive host that exhibited improved growth were chosen. We denote viruses from different lines and transfer numbers as follows: (1) $T7^{+,i}$—the phage with intact gene *1.3* from passage *i* of the preadaptation; and (2) $T7^{\Delta j,k}$—the phage from passage *k* of the *j*th line deleted for *1.3* ($j = 1, 2, 3$).

For the lines carrying deletions, the time at which the deletion was introduced is defined as $k = 0$. $T7^{\Delta 1}$ is the long-term line; $T7^{\Delta 2}$ and $T7^{\Delta 3}$ are short-term lines.

Results

Preadaptation to Culture Conditions

The ultimate goal was to adapt a virus to compensate for the sudden deletion of its gene for DNA ligase (gene *1.3*). The opportunity for compensatory adaptation was provided by growing the deletion mutant virus under a standard set of conditions. Changes that arose are likely to be strictly compensatory only to the extent that the parental virus was fully adapted to the culture conditions before the deletion was introduced. We thus adapted wild-type T7 ($T7^{+}$) on the strain IJ434 for 10 passages in liquid before introducing the *1.3* deletion. Fitness improved considerably during this 10-passage phase of preadaptation and reached what we thought was a likely maximum, based on our experience with the same virus on other hosts (a fitness of 6.4 doublings per generation for $T7^{+,10}$; fig. 1). It was later discovered that an even higher fitness could be obtained (7.6 for $T7^{+,20}$). The incomplete preadaptation with this line might be thought to preclude its use for compensatory evolution on the grounds that noncompensatory changes might continue to evolve. However, we developed a direct method of discriminating compensatory from noncompensatory changes, which allowed use of $T7^{+,10}$ in studying compensatory evolution (see below).

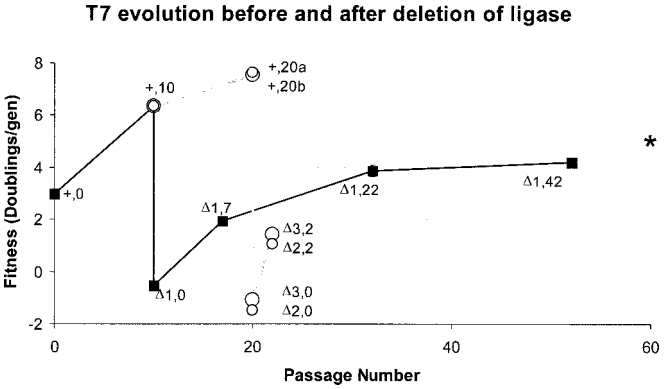

FIG. 1.—Evolutionary trajectories of T7 on IJ434. Labels near each data point are the superscripts of the phages analyzed at that point (e.g., +,10 is from $T7^{+,10}$). The deletion of the ligase gene (*1.3*) is marked by an immediate and profound drop to negative fitness values. Three lines of the deletion mutant are shown, two of them overlapping closely. Standard error bars (±1 SE) are given for all points, but in most cases they are obscured by the data point. The asterisk is at a fitness of 5.0, observed for one line derived from $T7^{\Delta 1,42}$ and is the highest fitness we have observed for a *1.3*− phage.

Deletion of Ligase is Profoundly Deleterious

On a ligase-proficient host, T7 gene *1.3* is not essential because the host ligase compensates. Compensation is not complete; gene *1.3* mutants are reduced in fitness by about 2 units (equivalent to a drop in burst size of about fourfold, relative to the $T7^{+}$), but the virus still forms large plaques at a relative efficiency of 1. The host chosen for the present study (IJ434) carries a ligase mutation (Studier 1973) that substantially reduces enzyme activity. $T7^{+}$ grows normally on this host, but any T7 mutant that lacks gene *1.3* activity is seriously debilitated. Immediately after the *1.3* deletion was introduced into the virus, fitness on IJ434 dropped from 6.4 (for $T7^{+,10}$) to −0.5 (for $T7^{\Delta 1,0}$) (fig. 1). Slightly greater drops in fitness of two short-term lines, $T7^{\Delta 2,0}$ $T7^{\Delta 3,0}$ are described below. The fitness of 6.4 is equivalent to a single phage finding and adsorbing to a host, replicating, and releasing 81 progeny, all in a 15-min interval; a fitness of −0.5 represents a burst of less than one phage during this same interval. Note that these values are not empirical burst sizes; they are merely alternative representations of the fitness values.

Major Recovery Occurs

A substantial portion of the fitness that was lost because of the introduction of the gene *1.3* deletion was recovered on subsequent passaging (fig. 1). The $T7^{\Delta 1}$ line was carried through 7 passages on plates and a further 35 passages in flasks, for a final fitness of 4.2 for the phage $T7^{\Delta 1,42}$. The fitness profile across transfers suggests that the final fitness from long-term recovery will remain substantially below that of the intact phage. Indeed, a fitness of only 5.0 was obtained after 40 transfers of a stock derived from $T7^{\Delta 1,42}$ and another *1.3*−

Table 1
Nucleotide Changes During Recovery After Deletion of T7 Ligase (gene *1.3*)

Base	Change	Gene	Gene Function/Effect of Mutation	$T7^{\Delta 1,0}$	$T7^{\Delta 1,7}$	$T7^{\Delta 1,15}$	$T7^{\Delta 1,42}$	$T7^{\Delta 2,0}$ $T7^{\Delta 3,0}$	$T7^{\Delta 2,2}$	$T7^{\Delta 3,2}$
2882*	C → T R288K	*0.7*	Protein kinase; stops host transcription	−	+	+	+	+	(+)	(+)
7781–7898	Deletion	*1.5*	Unknown	−	+	+	+			
10131–10132	G insert	*2.8*	Polar on gene 3	−	+	+	+	−	+	−
10332	C → T Q26X	*3*	Endonuclease	−	(−)	−	−	−	−	+
11929	A → G K122R	*4*	Primase-helicase	−	+	+	+	−	−	
15091	G → A A247T	*5*	DNA polymerase	−	+	+	+	−	−	
18518	G → A	Intergenic	None obvious	−	−	−	+	−	−	
36241*	A → G D540N	*17*	Tail fiber	−	−	+	+	+	(+)	(+)
36374*	C → T I11V	*17.5*	Holin (lysis)	−	−	+	+	+	(+)	(+)
36399*	C → T T19M	*17.5*	Holin (lysis)	−	−	−	+	+	(+)	(+)
37141*	G → T R75S,G37V	*18.5, 18.7*	Lysis	−	−	−	+	+	(+)	(+)

NOTES.—Substitutions marked with an asterisk (*) are not specifically compensatory for the ligase deletion. The first four T7 columns list the status of each change at different stages of the long-term recovery; the last three columns are for the two short-term recoveries. A blank cell means that the change was not assayed. Parentheses ( ) indicate an inferred state, based on ancestral states.

phage (data not shown). Over all lines and individual isolates, the fitness evolved from a low of −1.5 when the deletion was first introduced to a maximum of 5.0; in contrast, the highest fitness of a *1.3*+ phage was 7.6 (see below). On this log scale of fitness, therefore, approximately 70% of the possible fitness was recovered by compensatory evolution. On an absolute scale, the fitness of the recovered virus is only 17% that of the maximum fitness of the intact virus (from a low of 0.2%).

Not All Mutations Are Compensatory for the Loss of Ligase

In order to identify the mutations that occurred during adaptation of the gene *1.3* ligase deletion mutant, the complete genome sequence of a purified phage isolate from $T7^{\Delta 1,42}$ was obtained and compared to the published sequence of $T7^{+}$ (Dunn and Studier 1983). This comparison identified all potential sites of T7 evolution. Sequences were then obtained from $T7^{+,10}$ at each of these sites to identify which changes occurred after the loss of gene *1.3*. Ten changes were detected (table 1). The status of the 10 mutations was also assayed at two different times (passage 7 and 15) of the recovery phase to determine when they arose (table 1).

Despite the 10-passage interval of preadaptation to IJ434, any mutation that arose during the recovery phase could merely enhance growth of any phage on IJ434, instead of specifically compensating for the deletion. This explanation remains a formal possibility, no matter how extensive the preadaptation phase, so we developed a test to determine whether substitutions are compensatory. In this method, the mutations that fixed in the evolved line are introduced (by recombination) into the parental phage background ($T7^{+,10}$) at moderate frequency. On subsequent passage, any mutations that benefit the ancestral genome (and are thus not compensatory) will sweep through the population and distinguish themselves from those that ascend only in the presence of the deletion. The advantage of this method is that every mutation which evolves during a recovery can be introduced into the ancestral background at a high enough frequency that it will quickly ascend if it is beneficial in that background.

This manipulation is facilitated by the fact that T7 recombines at a high rate. Infecting $T7^{+,10}$ and $T7^{\Delta 1,42}$ into the same host allows recombination between the two phages, and serially passaging the heterogeneous phage lysates on IJ434 selects for recovery of the fittest phage. Because fitness of a *1.3* deletion mutant is substantially less than that of $T7^{+}$, 1.3^{+} phages will dominate lysates obtained on IJ434. Any of the 10 mutations found in $T7^{\Delta 1,42}$ that recombines into and evolves in the 1.3^{+} background can then be interpreted as benefiting the parental virus and thus not compensatory for the loss of *1.3* activity. When multiple mutations are evaluated simultaneously, this method does rely on T7 recombination to produce the full array of genotypes with different combinations of mutations; the extent to which this goal is accomplished in any trial is largely unknown.

Two independent lines were created in this fashion and passaged to identify noncompensatory mutations. $T7^{+,10}$ and $T7^{\Delta 1,42}$ were mixed and infected into IJ434 at high multiplicity. The lysate was passaged further on IJ434 at high multiplicity for three cycles, the fourth lysate was then mixed in equal proportions with $T7^{+,10}$ and $T7^{\Delta 1,42}$ and grown for another six cycles, giving $T7^{+,20a}$ and $T7^{+,20b}$. The fitnesses of both the lysates were determined to be 7.6, a significant increase over $T7^{+,10}$ (6.4). Thus, the fitness against which recovery of the *1.3* deletion mutant, $T7^{\Delta 1,42}$, should be compared is 7.6 rather than 6.4. In other words, some of the recovered fitness of $T7^{\Delta 1,42}$ can be attributed to additional adaptation of $T7^{+}$ to IJ434, instead of a response to the deletion of gene *1.3*. Both $T7^{+,20a}$ and $T7^{+,20b}$ contained the four late-region changes shown in table 1 (genes *17*, *17.5*, and *18.5*/*18.7*), and $T7^{+,20a}$ also contained the gene *0.7* mutation. It seems likely that the four late-region mutations are not compensatory for loss of ligase activity, and this analysis also questions the compensatory nature of the *0.7* change. The latter cannot offer a large

benefit to the parental phage, but the fact that it was found in one culture suggests that it might provide a small fitness benefit. This conclusion is reinforced by the following observation. The gene *1.3* deletion was introduced into $T7^{+,20b}$ by phage-plasmid recombination. The particular recombinant phage selected for use in a short-term recovery experiment ($T7^{\Delta 3,0}$, see below) was sequenced and found to contain the *0.7* mutation. Thus although not a major constituent, phages containing the *0.7* mutation were also present in the $T7^{+,20b}$ stock, so this mutation can also be considered to have been caused by an adaptation of the parental phage to growth on IJ4343 and not as compensating for the *1.3* deletion.

Five changes thus remain as candidate compensatory mutations for the loss of ligase: a deletion of gene *1.5*, a frameshift mutation in gene *2.8*, a missense mutation in each of the genes *4* and *5*, and an intergenic substitution upstream of gene *6.5*. Genes *4* and *5* have known functions in DNA metabolism: primase-helicase and DNA polymerase, respectively. The precise function of gene *2.8* has not been determined, but sequence comparisons have suggested that gene *2.8* shows similarity to homing endonucleases (Gorbalenya 1994; Shub and Goodrich-Blair 1994; Dalgaard et al. 1997). However, it will be argued below that the frameshift mutation is polar on gene *3* expression and that the latter is the important factor in compensating for loss of DNA ligase activity. Gene *3* codes for an endonuclease that degrades the host chromosomal DNA and resolves Holliday junctions during T7 recombination. Thus 3 of the 5 compensatory mutations affect DNA metabolism, where ligase also functions. The function of gene *1.5* is unknown, and there is no effect that can be inferred for the substitution in the intergenic region between gene *6.3* and the *6.5* promoter.

Two Short-term Recoveries

Deletions of *1.3* were introduced into $T7^{+,20a}$ and $T7^{+,20b}$, yielding $T7^{\Delta 2,0}$ and $T7^{\Delta 3,0}$, respectively. $T7^{\Delta 2,0}$ and $T7^{\Delta 3,0}$ formed plaques on IJ434 at an efficiency of less than 10^{-4}. A single plaque on IJ434 derived from each stock was then purified on IJ434 to obtain second-generation stocks ($T7^{\Delta 2,2}$ and $T7^{\Delta 3,2}$) that could be used for genetic and fitness analyses. Fitness was improved considerably in both the isolates but remained well below that of the long-term recovery line $T7^{\Delta 1,42}$ (fig. 1). However, fitnesses in the isolates are not easily accommodated with fitnesses in the long-term line, and these comparisons suggest that strong fitness interactions underlie the mutational effects. For example, although $T7^{+,20a}$ and $T7^{+,20b}$ each contained five of the mutations from $T7^{\Delta 1,42}$ whose combined effect in that background was at least +2 fitness units (from fig. 1 and table 1), the fitnesses after deletion of *1.3* ($T7^{\Delta 2,0}$ and $T7^{\Delta 3,0}$) were each lower than that of $T7^{\Delta 1,0}$, significantly lower for $T7^{\Delta 2,0}$. This result is opposite of what is expected if the fitness effects are additive or multiplicative. A further suggestion of interactions is that the fitnesses of isolates $T7^{\Delta 2,2}$ and $T7^{\Delta 3,2}$ were as much improved over their $T7^{\Delta 2,0}$ and $T7^{\Delta 3,0}$ ancestors as was $T7^{\Delta 1,7}$ over its $T7^{\Delta 1,0}$ ancestor. Table 1 reveals that $T7^{\Delta 1,7}$ acquired five mutations from its ancestor, whereas we will suggest below that $T7^{\Delta 2,2}$ and $T7^{\Delta 3,2}$ each had possibly only one mutation over their ancestors. Thus, one mutation in each of $T7^{\Delta 2,0}$ and $T7^{\Delta 3,0}$ had the same benefit as five in $T7^{\Delta 1,0}$.

By exchanging fragments of $T7^{\Delta 2,2}$ with $T7^{\Delta 2,0}$, the region of the $T7^{\Delta 2,2}$ genome responsible for its increased plating efficiency on IJ434 was located between nucleotides 9489 and 20065 (*Mlu* I—*BstE* II fragment) of $T7^{+}$ DNA. The sequence of this fragment was found to contain a G insert at nucleotide #10131 as the only mutation. This mutation lies in gene *2.8* and is the same one found in the $T7^{\Delta 1}$ line.

Unlike $T7^{\Delta 2,2}$, the isolate $T7^{\Delta 3,2}$ plated better on the ligase-defective host IJ434 than on its ancestral parent *E. coli* B, which is ligase-proficient. This result suggested that the two phages might have compensated for their lack of ligase activity in different ways. The phenotype of $T7^{\Delta 3,2}$ was found to be similar to that of T7 gene *1.3,3* double amber mutants (Sadowski 1974), who suggested that the loss of essential activity of the gene *3* endonuclease could be alleviated by a reduction in ligase activity in the infected cell. Consequently, we predicted that the mutation in $T7^{\Delta 3,2}$ that compensated for the deletion of gene *1.3* would also lie in gene *3*. The sequence of gene *3* of $T7^{\Delta 3,2}$ was therefore determined, and a mutation converting codon 26 into an amber stop codon was found. A fragment of the $T7^{\Delta 3,2}$ genome containing this mutation was cloned into a plasmid and transformed into the amber-suppressing host IJ511. This plasmid-containing strain is fully permissive for $T7^{\Delta 3,2}$ and was used to recombine the amber mutation back into $T7^{\Delta 3,0}$. Two recombinants were examined, both plated on IJ434 orders of magnitude more efficiently than $T7^{\Delta 3,0}$, but the two recombinants appeared different. One recombinant plated 10-fold better than the other, and the two phages gave different plaque morphologies. Thus, whereas these data show that loss of gene *3* activity allows a T7 ligase deletion mutant to grow on the ligase-deficient host IJ434, other mutations that augment the effects of the gene *3* amber mutation could either have been lost during the recombination or could have been selected readily during plaque growth.

The Mutation in Gene *2.8* is Polar on Gene *3*

The frameshift mutation in $T7^{\Delta 1,42}$ and $T7^{\Delta 2,2}$ gene *2.8* generates a stop codon almost immediately in the downstream sequence, resulting in the synthesis of a truncated gp2.8 (93 of 139 residues). It was thus possible that compensation for the ligase deletion in $T7^{\Delta 1,0}$ and $T7^{\Delta 2,0}$ occurred by a complete loss of or by a change in gp2.8 activity. To determine whether loss of activity was important we constructed a precise deletion of gene *2.8* (#9857–10257 of $T7^{+}$ DNA) on a plasmid containing parts of genes *2.5* and *3*. Because the ribosome-binding site of gene *3* is within the coding region of gene *2.8*, the deletion was designed to allow gene *3* expression from the gene *2.8* ribosome-binding site. The plasmid was recombined with $T7^{\Delta 2,0}$, and the desired

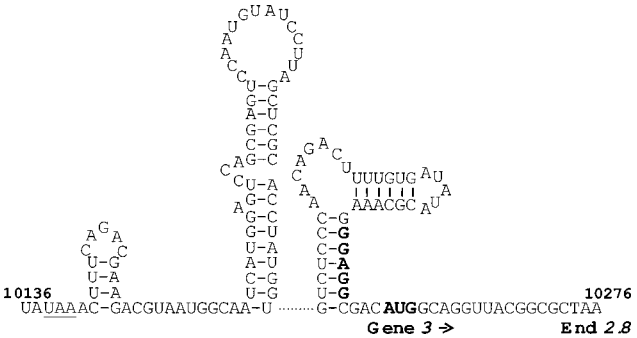

FIG. 2.—Predicted minimum free energy secondary structure of gene *2.8*-gene *3* RNA near the gene *3* initiation codon. Numbers correspond to T7 DNA sequence coordinates. The 5′ end of the sequence corresponds to the first base following the partial deletion of gene *2.8;* the gene *2.8* termination codon is the 3′ end of the sequence. The frameshift mutation found in $T7^{\Delta1,42}$ and $T7^{\Delta2,2}$ would lead to termination of gene *2.8* translation at three bases underlined (positions 10138–10140). The ribosome-binding site and the initiation codon of gene *3* are shown in bold face.

recombinant phages isolated after screening for the deletion by PCR. Recombinants were then tested for growth on the ligase-deficient host IJ434 and found to be comparable to the parent phage $T7^{\Delta2,0}$. Thus loss of gene *2.8* activity is not compensatory for the ligase deletion in $T7^{\Delta2,0}$.

It remained possible that an altered activity of gp2.8 was important in compensating for the loss of ligase activity. However, a different explanation is possible that could explain the effects of the frameshift mutation. Premature termination during translation of the frameshifted gene *2.8* could affect expression of gene *3.* The ribosome-binding site for gene *3* lies within the coding sequences of gene *2.8* (Dunn and Studier 1983). A minimum free energy folding of the RNA sequence upstream of the start codon of gene *3* reveals that the ribosome-binding site for gene *3* may be in duplex form and thus may be sequestered from ribosomes (fig. 2). If this happens normally in vivo, efficient translation of gene *3* would likely require translation of gene *2.8* RNA to make the gene *3* ribosome-binding site single stranded or to allow translational coupling. In the absence of translation of the gene *2.8,* translation of the gene *3* RNA would be reduced and would result in lower levels of gp3 endonuclease activity.

A second deletion of gene *2.8* was therefore constructed on a plasmid and recombined into $T7^{\Delta2,0}$ using a ligase-proficient host to avoid selection of compensatory mutations. Recombinants were screened by PCR for the presence of the deletion. This second deletion removed only those coding sequences of gene *2.8* (#9857–10135, including the initiation codon) that are translated in the frameshift mutant $T7^{\Delta2,2}$. It cannot therefore direct the synthesis of a gp2.8 with a hypothetical altered nuclease activity. However, this deletion mutant phage does leave intact the 3′ end of gene *2.8* that contains the gene *3* ribosome-binding site and the sequences predicted to base pair with it. The eop of $T7^{\Delta2,0}$, containing the deletion, on the ligase-deficient host IJ434 was at least three orders of magnitude higher than that of $T7^{\Delta2,0}$ ($P < 0.001$, *t*-test). This is comparable to the eop of the evolved phages $T7^{\Delta1,42}$ and $T7^{\Delta2,2}$ that contain the gene *2.8* frameshift mutation.

Both deletions effectively eliminate gene *2.8,* so the different phenotypes of each deletion are attributable only to their different regulatory effects on *3.* We therefore conclude that the important consequence of the gene *2.8* frameshift mutation of $T7^{\Delta1,42}$ and $T7^{\Delta2,2}$ in compensating for lack of DNA ligase activity is not the elimination of gp2.8 activity but is the prevention of gene *2.8* translation that in turn reduces translation of gene *3.* Thus the gene *3* amber mutation found in $T7^{\Delta3,2}$ and the gene *2.8* frameshift found in $T7^{\Delta1,42}$ and $T7^{\Delta2,2}$ both apparently increase fitness of a phage with the ligase deletion by reducing expression of gene *3.*

Discussion

At its most basic level, this study is an analysis of compensatory evolution in response to a major genomic perturbation. The phenomenon of compensatory evolution has a prominence in evolutionary biology, both as a mechanism to overcome the general decay in fitness from the long-term accumulation of deleterious mutation and also as a mechanism to ameliorate the deleterious pleiotropic effects of drug resistance mutations (Burch and Chao 1999; Whitlock and Otto 1999; Bjorkman et al. 2000; Levin, Perrot, and Walker 2000; Moore, Rozen, and Lenski 2000; Poon and Otto 2000). The identification of mutations involved in long-term compensatory evolution, in conjuction with knowing the biochemical effect of the original deleterious mutation adds a dimension to this study that is unique in the work on this topic.

Our study was also motivated from a more mechanistic perspective, whether it is possible to predict or explain how a viral genome will respond to the loss of an important gene. Bacteriophage T7 is one of the best-understood parasitic organisms at the level of molecular genetics. Its genome sequence has been known for almost two decades, the physiological functions of all its essential genes are understood and many of these genes have been characterized biochemically. Several conditionally essential genes have also been characterized at both the genetic and biochemical level. Despite these accomplishments, functions have not yet been ascribed to over one-third of the 59 T7 genes (Molineux 1999). Furthermore, although most regulatory elements in the genome may have been recognized and their overall properties understood, we do not appreciate the impact of the loss of one or a few of those elements, especially those that are repeated throughout the genome. Nevertheless, sufficient information has accumulated on the T7 biology to allow the formulation of an experimental parameter-determined computer model that simulates most aspects of the entire developmental cycle of the phage (Endy et al. 2000). This simulation even enjoys limited success in being able to predict the consequences of major perturbations to the phage genome, although it is clear that many facets of T7 biology need to be explored and understood before we can claim that we un-

derstand how the T7 genome functions. The properties of a genome are far greater than the simple arithmetical sum of the properties of its constituent parts; individual genes and their products interact and distinct biochemical reactions interconnect in creating a network of metabolic pathways that is manifest as the overall phenotype of the organism.

Predicting Compensatory Evolution

The wealth of information on the biology of T7 thus makes it a prime candidate for studies aimed at understanding genome function. Could we predict that compensatory evolution would occur? Could we rationalize compensatory changes in terms of T7 biology? Alternatively, or perhaps in addition, can evolutionary studies themselves provide new information on T7 biology?

At the outset, our expectation from work in other viral systems was that the loss of ligase activity would remain so deleterious to T7 that recovery to high fitness would require the genome to acquire new sequences through recombination or gene duplication and to replace ligase function by divergence of those sequences. This hope was not realized, and compensatory evolution occurred through point mutations and a deletion. Of the five compensatory mutations from the long-term line, three affected—either directly or indirectly—genes involved in DNA metabolism: endonuclease, DNA polymerase, and primase-helicase. DNA metabolism is also the realm of biochemistry and physiology that involves DNA ligase. A fourth gene, gene *1.5*, whose function is unknown, was deleted in one line in response to loss of DNA ligase activity, suggesting that it may also have a role in some aspect of DNA metabolism. The fifth compensatory mutation was a substitution in an intergenic region that has no easily inferred effect. In one of the two short-term recoveries, a different compensatory mutation was identified that directly affected one of the above three DNA metabolism genes. Thus four from the total of six compensatory mutations identified affect three genes involved in DNA metabolism, and a fifth mutation (fourth gene) can potentially be added.

Compensation for the loss of DNA ligase activity by the loss of gene *3* endonuclease was anticipated from the observation that the lethality of a gene *3* amber mutation could be suppressed by a *1.3* amber mutation if the phage was grown on a nonamber-suppressing, ligase-deficient host (Sadowski 1974). Those double amber mutant phages were constructed without exposure to selective growth conditions and as judged by their plaque size under selective plating conditions, they had low fitness. Thus the other compensatory mutations we identified during long-term evolution likely enhance the growth of ligase-deficient, endonuclease-deficient phages. Sadowski (1974) suggested that because endonuclease and ligase provide opposing functions in phage DNA metabolism (the nuclease breaks phosphodiester bonds, ligase seals them), levels of the two enzymes are normally balanced. Nuclease activity is necessary in T7 development, in part to degrade the host chromosome and in part to resolve Holliday junctions in genetic recombination. Too much ligase, relative to endonuclease, may repair nicks in DNA as rapidly as they are created and prevent their being utilized effectively. In contrast, too little activity of ligase, relative to endonuclease, could allow degradation of the phage genome. This line of thought may also explain some of the other compensatory mutations observed in this study; we suggest that the activities of other phage enzymes important for DNA metabolism (primase-helicase, DNA polymerase, gp1.5) must also be adjusted to accommodate absence of normal ligase activity. In more general terms, these results caution against using combination drug therapy against different proteins in a single metabolic pathway in order to achieve viral suppression: inhibiting ligase and endonuclease together actually increases viral fitness over that achieved by inhibition of either gene product alone.

Noncompensatory Mutations

Ten mutations evolved in the long-term line. Five were shown to be beneficial even in the presence of the ligase gene (the ancestral background) and thus are not compensatory. This method of inferring the compensatory nature of mutations has been used before (Schrag, Perrot, and Levin 1997), and it is superior to the alternative practice of inferring compensatory mutations either by preadaptation of the strain to the culture conditions or by comparing magnitudes of fitness increase between control lines and lines subjected to compensatory evolution (Moore, Rozen, and Lenski 2000). The limitation of relying on preadaptation is that it does not ensure the evolution of all beneficial mutations. Furthermore, the introduction of a highly deleterious mutation in a line may cause a mutation that is only weakly beneficial in the preadapted, ancestral line to be highly beneficial in the low-fitness line (e.g., Bull, Badgett, and Wichman 2000).

Whereas our test showed that five mutations were not compensatory, it cannot show that the other five are compensatory, although we have presumed this to be the case. A formal alternative is that some of them are mildly deleterious and evolved by hitchhiking with beneficial mutations. Four mutations appeared during the same interval (between passage 0 and passage 7), and if two or more of these swept through in the same genome, one or more could have been mildly deleterious. This is the mechanism behind the evolution of mutator genes, which occasionally create beneficial mutations in other parts of their genomes, and the beneficial mutations boost the frequency of the mutator genes through hitchhiking (Taddei et al. 1997). Given that one of the candidate compensatory mutations of T7 lies in DNA polymerase, it would be interesting to know if that mutation acts as a mutator.

Contributions to T7 Biology

Experimental evolutionary studies can be a useful adjunct to classical molecular genetic approaches for understanding a genome. Hence, a potential role for gene

1.5 in DNA metabolism is suggested by its loss during compensatory evolution. In addition, polarity of gene *2.8* translation on gene *3* translation has not been recognized previously in T7 biology, although this type of polarity has been shown for gene *1.2* (Saito and Richardson 1981). Translation of gene *1.1* is necessary for ribosomes to access the gene *1.2* ribosome-binding site. The reason for translational polarity is the same in both the gene pairs; the ribosome-binding site of the downstream gene is both a part of the coding sequence of the upstream gene and also can be sequestered in duplex RNA. Thus translation of the upstream gene is necessary for translation of the downstream gene.

Different Outcomes in Other Systems

Compensatory evolution in response to deletions has been studied in other viruses. In a trial testing the feasibility of a vaccine against HIV, a monkey was infected with an SIV carrying a 12-bp deletion of *nef*. In series of evolutionary steps in vivo, this virus restored the deleted sequence to almost the original amino acid composition (Whatmore et al. 1995). Olsthoorn and van Duin (1996) evolved the RNA phage MS2 following partial deletion of two regulatory hairpins (19 bases). One pathway of recovery recreated both hairpins, albeit with some differences from the original base sequence. The other pathway resulted in a single hairpin from vestiges of the two original hairpins. Thus, in both cases, compensatory evolution was local and repaired the deletions to restore the original function. A different outcome was obtained in an in vitro study with HIV-1 carrying three deletions. Berkhout et al. (1999) observed recovery via duplication of a regulatory region, not by restoration of any of the deletions. Fitness recovery appeared complete by some assays, although the virus was not necessarily preadapted to the assay conditions, and a quantitative evaluation of the extent of recovery is not possible.

Our T7 study thus parallels the above studies in showing extraordinary recovery of fitness from a debilitating deletion. Although T7, as a DNA virus, has an intrinsic rate of mutation lower than that of RNA viruses and retroviruses (Drake et al. 1998), considerable recovery occurred. The nature of compensatory evolution was different here than in SIV and MS2, in that recovery occurred without repair of the deletion. Perhaps as a consequence, compensatory evolution to restore original fitness levels in the ligase deletion mutant was incomplete and reached a maximum of about one-sixth the level of the wild-type phage. This result suggests that it may indeed be possible to permanently cripple a virus with deletions as a means of attenuating a virus for vaccines, as suggested by Berkhout et al. (1999). Compensatory evolution may be inevitable but in the right circumstances it can be limited. A cautionary note: viral virulence may not be simply related to viral reproductive rate (fitness), so that using a serial passage approach to attenuate a virus may not necessarily lead to a suitable vaccine strain.

All viruses rely on host factors for their reproduction. Recent studies have explored the possibility of developing antiviral agents that inhibit host factors rather than viral proteins (Flapan 2000; Schang, Rosenberg, and Schaffer 2000). One potential benefit of such a strategy is that viral resistance to inhibitors of host factors may be limited because the inhibitor does not directly attack or interact with viral proteins. The present study may be seen as a test of that model: the use of a host with a defective ligase is functionally equivalent to inhibiting host ligase with a drug. The extensive fitness recovery that was obtained in this study indicates that drugs which block essential host factors are still vulnerable to viral evolution of resistance. At the same time, the failure to recover more than one-sixth of the original fitness leaves hope that strategies of inhibiting host factors may achieve useful and permanent levels of suppression.

Attempts to study genome organization and function necessarily begin with an evolutionarily static view of how a genome works, based on the interactions of the elements, as they exist. The ability of a genome to recover from the loss of an essential gene through point mutations and other gene losses indicates that it possesses an evolutionary robustness that may extend far beyond the dynamic properties that would be evident from just a static view. The observations here may indirectly help explain why many genes of a genome do not make obvious contributions to immediate fitness (Raamsdonk et al. 2001): the ability to adapt may be an important determinant of genome composition (Kirschner and Gerhart 1998).

Acknowledgments

We thank W. Studier for many of the strains used in this work and three anonymous reviewers for comments on the manuscript. This work was funded by grants NSF 9726902 and NIH GM 57756 to J.J.B.

LITERATURE CITED

BALL, L. A., C. R. PRINGLE, B. FLANAGAN, V. P. PEREPELITSA, and G. W. WERTZ. 1999. Phenotypic consequences of rearranging the P, M, and G genes of vesicular stomatitis virus. J. Virol. **73**:4705–4712.

BEAUCHAMP, B. B., and C. C. RICHARDSON. 1988. A unique deoxyguanosinetriphosphatase is responsible for the OptA1 phenotype of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **85**:2563–2567.

BERKHOUT, B., K. VERHOEF, J. L. B. VAN WAMEL, and N. K. T. BACK. 1999. Genetic instability of live, attenuated human immunodeficiency virus type 1 vaccine strains. J. Virol. **73**: 1138–1145.

BJORKMAN, J., I. NAGAEV, O. G. BERG, D. HUGHES, and D. I. ANDERSSON. 2000. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. Science **287**:1479–1482.

BROWN, P. O., and D. BOTSTEIN. 1999. Exploring the new world of the genome with DNA microarrays. Nat. Genet. **21**(Suppl.):33–37.

BULL, J. J., M. R. BADGETT, and H. A. WICHMAN. 2000. Big benefit mutations in a bacteriophage inhibited with heat. Mol. Biol. Evol. **17**:942–950.

BURCH, C. L., and L. CHAO. 1999. Evolution by small steps and rugged landscapes in the RNA virus phi6. Genetics **151**:921–927.

CHIEN, C. T., P. L. BARTEL, R. STERNGLANZ, and S. FIELDS. 1991. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proc. Natl. Acad. Sci. USA **88**:9578–9582.

DALGAARD, J. Z., A. J. KLAR, M. J. MOSER, W. R. HOLLEY, A. CHATTERJEE, and I. S. MIAN. 1997. Statistical modelling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. Nucleic Acids Res. **25**:4626–4638.

DRAKE, J. W., B. CHARLESWORTH, D. CHARLESWORTH, and J. F. CROW. 1998. Rates of spontaneous mutation. Genetics **148**:1667–1686.

DUNN, J. J., and F. W. STUDIER. 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of genetic elements. J. Mol. Biol. **166**:477–535.

ENDY, D., L. YOU, J. YIN, and I. J. MOLINEUX. 2000. Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. Proc. Natl. Acad. Sci. USA **97**:5375–5380.

FLAPAN, D. 2000. Novel antiviral drug concept targets number of human viruses. Medscape Wire April 20 (www.medscape.com/medscapewire/2000/0400/medwire.0420.novel.html).

GOEBL, M. G., and T. D. PETES. 1996. Most of the yeast genomic sequences are not essential for cell growth and division. Cell **46**:983–992.

GORBALENYA, A. E. 1994. Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. Protein Sci. **3**:1117–1120.

JACKSON, R. J., A. J. RAMSAY, C. D. CHRISTENSEN, S. BEATON, D. F. HALL, and I. A. RAMSHAW. 2001. Expression of mouse interleukin-4 by a recombinant ectromelia virus suppresses cytolytic lymphocyte responses and overcomes genetic resistance to mousepox. J. Virol. **75**:1205–1210.

JEONG, H., S. P. MASON, A.-L. BARABASI, and Z. N. OLTVAI. 2001. Lethality and centrality in protein networks. Nature **411**:41–42.

KIRSCHNER, M., and J. GERHART. 1998. Evolvability. Proc. Natl. Acad. Sci. USA **95**:8420–8842.

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.

MOLINEUX, I. J. 1999. T7 bacteriophages. Pp. 2495–2507 *in* T. E. Creighton, ed. Encyclopedia of molecular biology. Wiley, New York.

MOORE, F. B., D. E. ROZEN, and R. E. LENSKI. 2000. Pervasive compensatory adaptation in *Escherichia coli*. Proc. R. Soc. Lond. B **267**:515–522.

OLSTHOORN, R. C. L., and J. VAN DUIN. 1996. Evolutionary reconstruction of a hairpin deleted from the genome of an RNA virus. Proc. Natl. Acad. Sci. USA **93**:12256–12261.

PETERSON, S. N., and C. M. FRASER. 2001. The complexity of simplicity. Genome Biol. **2**:2002.1–2002.8.

POON, A., and S. P. OTTO. 2000. Compensating for our load of mutations: freezing the meltdown of small populations. Evolution Int. J. Org. Evolution **54**:1467–1479.

RAAMSDONK, L. M., B. TEUSINK, D. BROADHURST et al. (13 co-authors). 2001. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat. Biotechnol. **19**:45–50.

SADOWSKI, P. D. 1974. Suppression of a mutation in gene *3* of bacteriophage T7 (T7 endonuclease I) by mutations in phage and host polynucleotide ligase. J. Virol. **13**:226–229.

SAITO, H., and C. C. RICHARDSON. 1981. Processing of mRNA by ribonuclease III regulates expression of gene *1.2* of bacteriophage T7. Cell **27**:533–542.

SCHANG, L. M., A. ROSENBERG, and P. A. SCHAFFER. 2000. Roscovitine, a specific inhibitor of cellular cyclin-dependent kinases, inhibits herpes simplex virus DNA synthesis in the presence of viral early proteins. J. Virol. **74**:2107–2120.

SCHRAG, S. J., V. PERROT, and B. R. LEVIN. 1997. Adaptation to the fitness costs of antibiotic resistance in *Escherichia coli*. Proc. R. Soc. Lond. B **264**:1287–1291.

SHUB, D. A., and H. GOODRICH-BLAIR. 1994. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. TIBS **19**:402–404.

STUDIER, F. W. 1969. The genetics and physiology of bacteriophage T7. Virology **39**:562–574.

STUDIER, F. W. 1973. Genetic analysis of non-essential bacteriophage T7 genes. J. Mol. Biol. **79**:227–236.

TADDEI, F., M. RADMAN, J. MAYNARD SMITH, B. TOUPANCE, P. H. GOUYON, and B. GODELLE. 1997. Role of mutator alleles in adaptive evolution. Nature **387**:700–702.

TURNER, P. E., and L. CHAO. 1999. Prisoner's dilemma in an RNA virus. Nature **398**:441–443.

WAGNER, A. 2000. Robustness against mutations in genetic networks of yeast. Nat. Genet. **24**:355–361.

WHATMORE, A. M., N. COOK, G. A. HALL, S. SHARPE, E. W. RUD, and M. P. CRANAGE. 1995. Repair and evolution of *nef* in vivo modulates simian immunodeficiency virus virulence. J. Virol. **69**:5117–5123.

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.

KEITH CRANDALL, reviewing editor

Accepted October 4, 2001