

# Genic Incompatibilities in Two Hybrid Bacteriophages

Darin R. Rokyta\* and Holly A. Wichman†

\*Department of Biological Science, Florida State University; and †Department of Biological Sciences, University of Idaho

Horizontal gene transfer and recombination play a major role in microbial evolution and have been detected in diverse groups, including many of medical relevance such as HIV and dengue virus. In the absence of mechanistic barriers, the evolutionary success of a particular recombination event is determined by whether the recombinant genotype suffers a fitness cost through the disruption of favorable epistatic interactions within the genome, and if so, the extent to which this fitness cost might be mitigated by subsequent compensatory evolution. To investigate the importance of epistatic interactions between genes and the evolutionary viability of a homologous recombination event between diverged ancestral genotypes, we constructed two recombinant microvirid bacteriophages by exchanging their alleles of the gene encoding the coat protein. The coding sequences for this gene differ by approximately 8% at the amino acid level and were interchanged between two ancestral phages related to  $\phi$ X174 and well adapted to their culture conditions. Because the recombinant phages showed drastically reduced fitnesses, we further explored their evolutionary viability by subjecting replicate lines of each of them to selection. We found that all four lineages achieved fitnesses commensurate with ancestral fitnesses in as few as 60 generations, and on average, the first substitution accounted for more than half of the total fitness recovery. Fitness recovery required three to five substitutions in each lineage, and overall eight of the nine essential phage genes were involved, suggesting extensive epistatic interactions throughout the genome. Interestingly, the proteins with the most extensive and apparent physical interactions with the exchanged protein in the viral capsid did not appear to have much of a role in fitness recovery. This result appears to be a consequence of the conservation of the amino acid residues involved in the interactions. It suggests that strong epistatic interactions are less important than weaker, transient ones in producing genic incompatibilities because they preclude variability in the interacting regions of the proteins.

## Introduction

Horizontal gene transfer and recombination are fundamental components of microbial adaptation and speciation (de la Cruz and Davies 2000), and the extensive statistical evidence of homologous recombination in microbes attests to its importance in their evolution. For example, homologous recombination has been detected in hepatitis C virus (Yun et al. 1996; Colina et al. 2004; Sentandreu et al. 2008), dengue virus (Holmes et al. 1999), and HIV (e.g. Sabino et al. 1994; Leitner et al. 1995; Robertson et al. 1995; Blackard et al. 2002; Charpentier et al. 2006). A number of potential functional barriers to homologous recombination can potentially limit the range of recombinants produced (Roberts and Cohan 1993; Zawadzki et al. 1995; Vulić et al. 1997; Majewski and Cohan 1998, 1999), but just as importantly, selective barriers may also act. Although recombination can allow shuffling of existing variation and much larger genetic changes than simple point mutation, epistasis may render the products of recombination unfit.

Epistatic interactions have long been of interest to evolutionary biologists because their precise form influences a variety of phenomena. For example, epistasis is thought to be important in the evolution of sexual reproduction (Otto and Feldman 1997) and as a mechanism facilitating speciation (Dobzhansky 1936; Muller 1939; Orr 1995; Coyne and Orr 2004; Good et al. 2007). From a theoretical perspective, interactions between genes and between sites within genes determine the “ruggedness”

of the fitness landscape (Kauffman and Levin 1987; Kauffman 1993; Weinreich et al. 2005), and the nature and extent of these interactions are the major parameters in models of fitness landscapes (Kauffman and Levin 1987; Kauffman 1993; Perelson and Macken 1995; Orr 2006).

Compensatory evolution is a manifestation of a particular form of epistasis referred to as sign epistasis (Poon and Chao 2005; Weinreich et al. 2005). A compensatory mutation is defined as being beneficial only in the context of some form of deleterious mutation and is thus an example of a mutation whose “sign” (beneficial or deleterious) is determined by the genetic context in which it arises. Compensatory mutations thus provide information on ruggedness of the local fitness landscape (Poon and Chao 2005) and can relieve the loss of fitness due to accumulation of deleterious mutations (Poon and Otto 2000). Compensatory mutations can be either within the same gene as the original deleterious mutation or in another gene (Poon et al. 2005). The latter can provide information on epistatic interactions between genes and identify proteins in the same biochemical pathway or that physically interact with the debilitated protein. Of course, the deleterious mutation need not be a simple point mutation (e.g. Poon and Chao 2005); it could also be, among other things, a large number of such mutations (Bull et al. 2003) or even a deletion (Rokyta et al. 2002).

In the study reported here, we first examined the importance of epistatic interactions between genes and then used disrupted intergenic epistatic interactions as a basis for subsequent compensatory evolution. In particular, we examined the fitness consequences of homologous recombination between two single-stranded DNA microvirid bacteriophages. The two initial phage genotypes, ID12 and ID2, were well adapted to our culturing conditions (Rokyta et al. in press), so both ancestors had high initial fitnesses. To focus on intergenic interactions, we exchanged the entire coding regions of a single gene, the phage coat

Key words: epistasis, recombination, horizontal gene transfer, bacteriophage, compensatory evolution.

E-mail: drokyta@bio.fsu.edu.

*Mol. Biol. Evol.* 26(12):2831–2839. 2009

doi:10.1093/molbev/msp199

Advance Access publication September 2, 2009

protein gene. We bypassed any functional barriers to recombination by artificially constructing chimeric bacteriophages to look at the nature of selective barriers imposed on the products of recombination. Any fitness reduction could be due only to the disruption of favorable intergenic interactions. By constructing these two recombinant phage genotypes and subsequently subjecting them to selective laboratory conditions, we addressed three main questions: 1) does recombination entail a fitness cost; 2) can fitness be recovered; and 3) what genes are involved in recovery?

## Materials and Methods

### Ancestral genotypes

The two microvirid bacteriophages used in this study, ID12 and ID2, were originally isolated and described by (Rokyta et al. 2006) and subsequently adapted to culture conditions by (Rokyta et al. in press). The GenBank accession numbers of the original isolates are DQ079905 (ID12) and DQ079890 (ID2). As our ID12 ancestor, we selected a random genetic isolate from the ID12a60 population of (Rokyta et al. in press). This isolate possessed the two mutations that fixed in the ID12a lineage of Rokyta et al. (in press): 1970 (A→G) and 4919 (G→A). In addition, the ID12 isolate had a mutation at position 4843 (A→G) that results in a K→E substitution at amino-acid position 110 in protein H. We sequenced three additional genetic isolates from this population and found that one had this mutation and two did not. Therefore, half of the sequenced isolates had this third mutation, suggesting that it was at appreciable frequency in the ID12a60 population and probably adaptive. All three mutations were nonsynonymous. Our ID2 ancestor was selected at random from the ID2a200 population of Rokyta et al. (in press) and matched the consensus sequence described by Rokyta et al. (in press) exactly. Relative to the published sequence of ID2, therefore, our ID2 isolate had the following nine mutations: 858 (A→G), 1570 (C→T), 1783 (C→T), 2087 (C→T), 3264 (G→A), 3932 (T→C), 4882 (C→T), 4902 (G→A), and 5070 (C→A). All these except for the mutation at position 2087 were nonsynonymous substitutions.

### Building hybrid phages

The coat protein gene, which is named F, was selected for exchange. Based on homology with other microvirid bacteriophages, the mature viral capsid contains 60 copies of the coat protein, and it makes up the majority of the capsid. Gene F occupies positions 2552–3835 in phage ID12 and positions 2546–3829 in phage ID2 and has been found to be important in determining host range (Crill et al. 2000) and in adaptation to high temperature (Kichler Holder and Bull 2001). A similar recombination event involving gene F of quite divergent phages has been hypothesized to have occurred in the history of the microvirid family (Rokyta et al. 2006).

Hybrid phages were constructed through a modified version of the site-directed mutagenesis protocol described by (Pepin et al. 2006). For each gene exchange, two sets of primers were designed for a total of eight primers. One

set was designed to amplify the allele of gene F from the “donor” phage and the other to amplify the genome without gene F from the “recipient” phage. Each primer was approximately half ID12 sequence and half ID2 sequence. Primers were 31–34 nucleotides in length, roughly centered on the start or stop codons of gene F, and their sequences matched the desired chimeric phage sequence. For each exchange, the two pieces of the genome were amplified separately in a standard polymerase chain reaction (PCR) (25 cycles). The two fragments were then purified and combined in equal copy numbers in another PCR reaction for 10 cycles without primers to fill-in the complete genome. The product of this reaction was purified, electroporated into *Escherichia coli* C, and plated. Phage from the resulting plaques were confirmed to be recombinant and free of additional mutations by full genome sequencing.

### Fitness assays and recovery

Fitness assays were performed exactly as described by (Rokyta et al. in press). Fitness was measured as the  $\log_2$  increase in total phage per hour. The phage host, *E. coli* C, was grown to a concentration of  $1\text{--}2 \times 10^8$  cells per milliliter in phage Lysogeny broth (10 g NaCl, 10 g tryptone, and 5 g yeast extract per liter) supplemented with 2 mM  $\text{CaCl}_2$  in 125-ml flasks at 37°C shaking at 200 rpm in an orbital water bath. Phage were added ( $\sim 10^4\text{--}10^5$ ) and grown for 40 min. Growth was terminated with chloroform. Fitness was measured in five replicates for each genotype. Fitness recovery was performed through serial flask transfers under conditions similar to fitness assays, except that the number of phage used to initiate each growth period was increased by approximately 10-fold. Each lineage was initiated from an independent genetic isolate. All statistical analyses of fitness values were performed with R (R Development Core Team 2006).

## Results

### Relationships between the phages

The bacteriophages ID12 and ID2 are quite similar, with the same complement of 11 genes and similar genome lengths: 5529 for ID12 and 5486 for ID2 (Rokyta et al. 2006). They differ at  $\sim 19\%$  of nucleotide sites over the whole genome, excluding gaps. At the amino acid level, the coat protein genes (F) are about 8% different (34 of 426 amino acid sites differ). The other gene products composing the mature phage virion and its scaffolding proteins differ at the amino acid level as follows: the spike protein (G), 59%; the pilot protein (H), 14%; the DNA-binding protein (J), 25%; the internal scaffolding protein (B), 18%; and the external scaffolding protein (D), 9%. All these numbers exclude the start and stop codons and gaps. Note the high divergence between the spike proteins. Rokyta et al. (2006) hypothesized that it is the result of a recombination event in the distant past. Of the 11 genes, five (A\*, B, C, F, and K) are of identical lengths in the two ancestral phages. ID2's proteins are shorter for genes A (one amino acid), D (one amino acid), E (12 amino acids), and H (two amino acids). ID2 has longer proteins for genes J (one amino acid) and G

**Table 1**  
**A Summary of the Fitness Recovery of the Hybrid Phages**

Lineage	$w_{\text{anc}}^a$	SE <sup>b</sup>	$\Delta w_{\text{rec}}^c$	Flasks	$\bar{N}_H^d$	$w_i^a$	SE <sup>b</sup>	$w_f^a$	SE <sup>b</sup>	$\Delta w_{\text{comp}}^c$
ID12-ID2Fa	24.3	0.30	-13.2	70	$8.7 \times 10^4$	11.1	0.31	21.9	0.35	10.8
ID12-ID2Fb	24.3	0.30	-13.4	115	$3.3 \times 10^5$	10.9	0.38	22.5	0.31	11.6
ID2-ID12Fa <sup>e</sup>	21.9	0.68	-8.9 <sup>e</sup>	80	$3.9 \times 10^4$	13.0 <sup>e</sup>	0.41	20.3	0.22	7.3 <sup>e</sup>
ID2-ID12Fb	21.9	0.68	-14.5	75	$3.4 \times 10^5$	7.4	0.55	23.3	0.30	15.9

SE, standard error.

<sup>a</sup>Fitness ( $w$ ) was measured as the  $\log_2$  increase in population size per hour. Subscripts “anc”, “i” and “f” designate ancestral, initial and final fitness, respectively.

<sup>b</sup>Based on five replicates.

<sup>c</sup>Change in fitness. Subscripts “rec” and “comp” indicate the change caused by the introduction of new allele of F and the change through compensatory evolution, respectively.

<sup>d</sup>Harmonic mean of population sizes over the course of fitness recovery.

<sup>e</sup>In the initial genotype for this lineage, a beneficial mutation had already become fixed (see text).

(two amino acids). In all that follows, amino acid positions will be based on homology with ID12 genes.

### The cost of recombination

The two phage ancestors, ID12 and ID2, had adapted to our culturing conditions during a study by Rokyta et al. (in press). Both were allowed to adapt until fitness plateaued for at least 20 growth periods. Our ancestral phages were therefore highly fit; ID12 and ID2 showed 24.3 and 21.9 doublings per hour, respectively (table 1). The corresponding figures before this adaptation were 20.8 and 8.7 doublings per hour (Rokyta et al. in press). Reciprocal exchange of alleles of the coat protein gene F was then performed, generating two new phage genotypes labeled ID12-ID2F, which consists of the ID12 genome with its allele of F replaced by the homologue from ID2, and ID2-ID12F, which consists of the ID2 genome with its allele of F replaced by the homologue from ID12. The fitnesses of both recombinants were well below the fitness of either ancestor ( $t$ -test, two-sided, unequal variance,  $P \ll 0.001$  for all four comparisons). Introduction of the ID2 allele of gene F reduced the fitness of the ID12 ancestor from 24.3 to 11.1, a fitness cost of 13.2 doublings per hour (table 1). This cost represents a 9,400-fold reduction in number of progeny produced per hour. Similarly, introduction of the ID12 allele of gene F into the ID2 genome reduced fitness from 21.9 to 7.4 doublings per hour, a fitness cost of 14.5 doublings per hour (table 1). This cost corresponds to a 23,000-fold reduction in number of progeny per hour. At a generation time of about 15 min for these phages, these fitness costs correspond to a 10-fold and 12-fold reduction in progeny per generation for ID12 and ID2, respectively. In both cases, growth rate was reduced to less than half of the parental growth rate. Because both ancestral phages are highly fit in this environment, both alleles of F are high-fitness alleles in their original genetic contexts. These fitness costs therefore reflect only the effects of disrupting favorable epistatic interactions in the ancestral genomes.

### Fitness recovery

Each of the two hybrid phage genotypes was selected through serial flask transfers to recover the fitness lost to recombination through compensatory mutations. Each

genotype was selected for increased growth rate in two replicate lineages, which we labeled “a” and “b” (e.g. ID12-ID2Fa and ID12-ID2Fb). The ancestral isolate for the ID2-ID12Fa lineage had a significantly higher initial fitness than the isolate used for the ID2-ID12Fb lineage ( $t$ -test, two-sided, unequal variance,  $P \ll 0.001$ ; fig. 1 and table 1). The isolate used to initiate the ID2-ID12Fa lineage had acquired a mutation at position 3198 relative to the expected sequence (table 3). This mutation was found to be polymorphic in the original stock from which the isolate was derived and was adaptive, as it increased fitness by 5.6 doublings per hour.

All four lineages of recombinant phages achieved fitnesses commensurate with at least one of the ancestral fitnesses (fig. 1). Only one, ID2-ID12b, reached a fitness as high as that of the highest fitness ancestor, ID12 ( $t$ -test, two-sided, unequal variance,  $P = 0.19$  for ID2-ID12b and  $P < 0.02$  for ID12-ID2Fa, ID12-ID2Fb, and ID2-ID12Fa, Bonferroni corrected for four comparisons), but none reach a final fitness significantly different from that of ID2 ( $t$ -test, two-sided, unequal variance,  $P > 0.25$  for all lineages, Bonferroni corrected for four comparisons). Therefore, if complete fitness recovery is defined as reaching the fitness of at least one of the ancestors, then all four lineages achieved it. Furthermore, this fitness recovery occurred rapidly, taking less than 115 flask growth periods, corresponding to approximately 350 generations at a generation time of  $\sim 12$ –15 min.

We can further address how quickly recovery was achieved by assessing how long it took to reach the fitness of the ancestor with the lowest fitness, ID2. For each of the four lineages, we measured the fitness of intermediate populations as well as the initial isolate and final population (table 2), specifically, those of the 10th and 20th flasks and every 20th flask thereafter (fig. 1). For the ID12-ID2a lineage, fitness remained significantly below that of ID2 until the 40th flask ( $t$ -test, one-sided, unequal variance;  $P < 0.001$  for flasks 0, 10, and 20;  $P > 0.07$  for flasks 40, 60, and 70; table 2). For the ID12-ID2Fb lineage, fitness also remained below ID2s until the 40th flask ( $t$ -test, one-sided, unequal variance;  $P < 0.02$  for flasks 0, 10, and 20;  $P > 0.10$  for flasks 40, 60, 80, 100, and 115; table 2). For the ID2-ID12Fa lineage, fitness remained below ID2’s fitness at all comparisons ( $t$ -test, one-sided, unequal variances;  $P < 0.04$  for flasks 0 and 10, 20, 40, 60, and 80; table 2). Finally, for the ID2-ID12Fb lineage, the

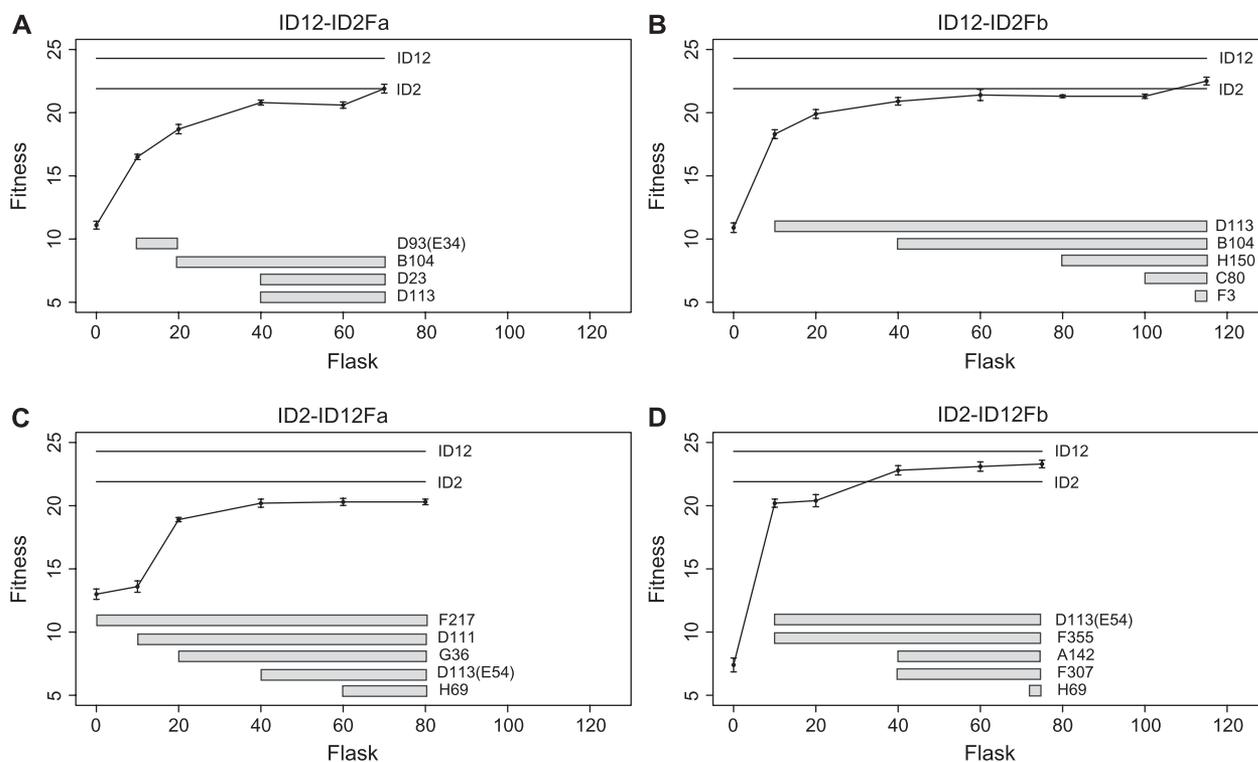


FIG. 1.—The trajectories of fitness recovery. The error bars indicate standard errors. The gray bars below the fitness trajectories indicate the presence of the mutations labeled at the right, on the basis of sequencing the populations for which fitness was measured. Mutations are labeled with the gene name and position of the affected amino acid residue. More information on these mutations is provided in table 3, and the raw fitness data from the populations are provided in table 2.

fitness of ID2 was reached by the 20th flask (*t*-test, one-sided, unequal variances;  $P < 0.05$  for flasks 0 and 10;  $P > 0.05$  for flasks 20, 40, 60, and 75; table 2). We found, therefore, assuming that the phages go through approximately three generations per flask growth period, that fitness can be recovered to ancestral levels within as few as 60 generations.

#### The genetics of fitness recovery

Fitness was recovered through only a handful of mutations for each of the four lineages (table 3). ID12-ID2Fb, ID2-ID12Fa, and ID2-ID12Fb recovered it through five substitutions, and ID12-ID2Fa recovered it in just three substitutions. Of the 18 observed substitutions, only 1 was silent. The microvirid phage genome encodes 11 gene products, labeled A, A\*, B, C, D, E, F, G, H, J, and K. Two of these (A\* and K) are not essential for replication and no compensatory mutations were found in either. Aside from the single-stranded DNA genome, the mature phage capsid consists of proteins F, G, H, and J. Four compensatory mutations occurred in the gene that was exchanged, gene F. A single substitution occurred in the spike protein (G), and three in the pilot protein (H). Eight substitutions altered the amino acid sequences of the internal and external scaffolding proteins (B and D, respectively). Finally, four mutations affected the coding sequences of genes involved in phage replication (A, E, and C). Note that, because of overlapping reading frames in the phage genomes, a single

mutation could potentially affect up to three genes. Overall, we find that changing the allele of the coat protein gene engendered nonsynonymous compensatory changes in eight of the nine essential genes, suggesting extensive epistatic interactions throughout the phage genome.

#### The timing of substitutions

In addition to the initial genetic isolate and final population of each lineage, we sequenced all the populations on which we measured fitness, as describe above (fig. 1). By sequencing whole populations, we can only determine when a particular mutation comes to high frequency. The order of fixation for some lineages could not be discerned entirely (fig. 1 and table 3).

All the early mutations (first or second) were either in the scaffolding genes, B and D, or the exchanged coat protein gene. Because these first mutations accounted for the majority of the total fitness improvement (see below), the conflict between the coat protein gene and the remainder of the genome probably arises at some stage in capsid assembly. All the substitutions in the other structural genes (G and H) and the genes involved in replication (A and C) occurred late in fitness recovery.

Two mutations were found to increase in frequency only to be lost again later. In the ID12-ID2Fa lineage, a mutation at position 2253 reached high frequency in the 10th flask growth period but was lost by the 20th (fig. 1). In the

**Table 2**  
**Fitness of Populations over the Course of Recovery**

Lineage	Flask	Fitness	SE <sup>a</sup>	P <sup>b</sup>
ID12-ID2Fa	0	11.1	0.31	<0.01
	10	16.5	0.21	<0.01
	20	18.7	0.37	<0.01
	40	20.8	0.20	0.09
	60	20.6	0.25	0.07
ID12-ID2Fb	70	21.9	0.35	0.53
	0	10.9	0.38	<0.01
	10	18.3	0.35	<0.01
	20	19.9	0.35	0.02
	40	20.9	0.30	0.11
	60	21.4	0.44	0.29
ID2-ID12Fa	80	21.3	0.11	0.21
	100	21.3	0.17	0.20
	115	22.5	0.31	0.78
	0	13.0	0.41	<0.01
	10	13.6	0.45	<0.01
	20	18.9	0.17	<0.01
ID2-ID12Fb	40	20.2	0.32	0.03
	60	20.3	0.28	0.04
	80	20.3	0.22	0.04
	0	7.4	0.55	<0.01
	10	20.2	0.32	0.03
	20	20.4	0.48	0.05
ID2-ID12Fb	40	22.8	0.37	0.87
	60	23.1	0.37	0.92
	75	23.3	0.30	0.94

SE, standard error.

<sup>a</sup>Based on five replicates.<sup>b</sup>Based on a one-sided *t*-test with unequal variances comparing the population fitness to the fitness of the ID2 ancestor with mean fitness 21.9 and SE 0.68 based on five replicate measurements.

ID12-ID2Fb lineage, the insertion of a C nucleotide in a poly-C tract in the J–F intergenic regions (positions 2529–2534) reached high frequency in flasks 80 and 100 but was subsequently lost (not shown in fig. 1). These two observa-

tions suggest the presence of some amount of clonal interference (Gerrish and Lenski 1998) in these lineages.

#### The first step in recovery

On average, each mutation increased fitness by 2.9 doublings per hour, including the mutation 3198 that we found to be in the isolate used to initiate the ID2-ID12Fa lineage by calculating the effect relative to the initial isolate of the ID2-ID12Fb lineage. In contrast, the first substitution increased fitness by about 6.8 doublings per hour on average. These are, of course, approximate values based on population fitnesses and population sequences. For the calculation, we assumed that the first substitution occurred by flask 20 for ID12-ID2Fa and by flask 10 for ID12-ID2Fb (fig. 1). For lineage ID2-ID12Fa, we assumed that the first mutation was at position 3198 and thus compared the fitness of ID2-ID12Fa0 with that of ID2-ID12Fb0. Finally, we assumed that two mutations became fixed by flask 10 in ID2-ID12Fb and that the two had equal effects. Nonetheless, the first step in compensatory evolution was clearly much larger than the average step (over twice as large). In fact, the first substitution accounted for more than half of the total fitness recovery on average.

#### Parallel and convergent evolution

For the two ID12-ID2F lineages, two mutations (1585 and 2312) occurred in parallel (table 3), so 50% of the mutations fixed in these two lines were parallel substitutions. Likewise, two mutations (2305 and 4684) occurred in parallel across the two ID2-ID12F lineages, so 40% of the mutations in these two lineages were parallel changes.

Because the exchanged alleles of gene F were more suited to their original genetic backgrounds, fitness

**Table 3**  
**A Summary of the Genetics of Fitness Recovery of Hybrid Phages**

Lineage	Nucleotide <sup>a</sup>	Change	Gene	Amino Acid <sup>c</sup>	Change	Order	Convergent	Parallel
ID12-ID2Fa	1585	A→G	B <sup>c</sup>	104	T→A	1	Yes	Yes
	2044	C→T	D	23	Silent	2 or 3	No	No
	2312	C→T	D <sup>d</sup>	113	R→C	2 or 3	No	Yes
ID12-ID2Fb	1585	A→G	B <sup>c</sup>	104	T→A	2	Yes	Yes
	1958	A→G	C	80	N→S	4	No	No
	2312	C→T	D <sup>d</sup>	113	R→C	1	No	Yes
	2562	T→C	F	3	V→A	5	No	No
	4964	C→T	H	150	T→I	3	No	No
ID2-ID12Fa	2298	A→G	D <sup>d</sup>	111	I→V	2	Yes	No
	2305	A→G	D(E)	113(54)	Q→R(S→G)	4	Yes	Yes
	3198 <sup>b</sup>	A→G	F	217	D→G	1	No	No
	4037	C→T	G	36	L→F	3	No	No
	4684	C→T	H	69	P→L	5	No	Yes
ID2-ID12Fb	482	A→G	A	142	I→V	3 or 4	No	No
	2305	A→G	D(E)	113(54)	Q→R(S→G)	1 or 2	Yes	Yes
	3468	A→G	F	307	N→S	3 or 4	Yes	No
	3611	C→T	F	355	P→S	1 or 2	No	No
	4684	C→T	H	69	P→L	5	No	Yes

<sup>a</sup>ID12 and ID2 differ in length, and therefore nucleotide positions are not directly comparable across these two genotypes.<sup>b</sup>Arose in plaque used to start the lineage.<sup>c</sup>Silent in gene A.<sup>d</sup>Silent in gene E.<sup>e</sup>Amino acid positions are based on the ID12 sequence.

recovery might reasonably be expected to make the new, “accepting” genomes more like the “donor” genomes. In other words, we might expect mutations involved in fitness recovery to converge on the original genomic context of the exchanged allele of F, or conversely for the exchanged allele of F to converge on the sequence of the original genotype. Indeed, this pattern was observed (table 3). Overall, 6 of the 18 mutations were convergent, including 1 mutation in the ID2-ID12Fb lineage (3468) in the new allele of F, which made it more like its ancestral F.

Mutations at amino acid position 113 in the external scaffolding gene D exhibited both parallel and convergent evolution (table 3). Each of the four lineages had substitutions at this position. The changes were all parallel across replicates. Both ID12-ID2F lineages changed from R to C, and both ID2-ID12F lineages changed from Q to R. The changes in the ID2-ID12F lineages converged on the ID12 ancestor, but the changes in the ID12-ID2F lineages were not convergent. A convergent change in these lineages (to Q) would have required two nucleotide substitutions, however, so the convergent amino acid substitution was not mutationally accessible.

## Discussion

By exchanging alleles of the coat protein gene between two microvirid genotypes, we demonstrated that recombination can entail a substantial fitness cost. Because both alleles were highly functional in their ancestral genetic backgrounds, this cost is clearly the result of epistatic interactions between genes. Although the reduction in fitness was dramatic, reducing growth rates to less than half of the parental rates, fitness recovery was complete and simple. All four lineages reached ancestral-level fitness within ~350 generations, and one lineage did so in only 60 generations. Furthermore, this fitness recovery required only three to five substitutions, and the first substitution accounted for the majority of fitness improvement. Despite the small number of compensatory mutations, they were found to affect eight of the eleven microvirid genes, including eight of nine essential genes. This result suggests that epistatic interactions are widespread throughout the genome and that the conflict between the new allele of gene F and the genome can be mitigated through multiple mechanisms.

Even if most genes in the microvirid genome are involved in epistatic interactions with the coat protein gene, the high rate of both parallel and convergent evolution suggests that only a small number of amino acid sites are involved. Across replicate recovery lineages, 40–50% of substitutions occurred in parallel, suggesting either that few sites are available for compensatory changes or that few sites have large effects. Furthermore, 6 of the 18 observed substitutions converged on the genetic background that provided the allele of the coat protein gene. The hybrid phages therefore evolved to become more similar to the ancestral donor genotypes, suggesting that the interactions between the genes involve not only specific sites but also specific amino acid residues at those sites.

The nature of the disrupted epistatic interactions between the genes in our hybrid microvirid genotypes can

be better understood by examination of the identities of compensatory mutations in relation to the extensive structural and experimental information available for microvirid bacteriophages (e.g. McKenna et al. 1992, 1994, 1996; Dokland et al. 1997; Bernal et al. 2003; Dokland et al. 1999; Novak and Fane 2004). For example, we know where and how the coat protein interacts with two of the three other structural proteins (G and J, but not H), and the interactions between the coat protein and the two scaffolding proteins (B and D) are also well characterized. These regions of interaction are natural places to expect to see compensatory evolution in response to a new allele of the coat protein gene.

## Compensatory evolution in the scaffolding proteins

Many of the compensatory mutations, including the majority of those that occurred early, affected the two scaffolding genes B and D. In fact, one lineage (ID12-ID2Fa) recovered entirely through mutations in these two genes (table 3), suggesting that morphogenesis is disrupted in the hybrids. Both lineages of ID12-ID2 acquired the same mutation at amino acid position 104 in the internal scaffolding protein encoded by gene B. The internal scaffolding protein is 120 residues long, and its C-terminus (~61–120) is known to determine coat protein specificity (Burch and Fane 2000) and contains the sites known to be contact points between the coat and internal scaffolding proteins (Dokland et al. 1999). In complementation experiments, internal scaffolding proteins have been found to be cross-functional across genetic distances greater than that between the two genotypes ID12 and ID2, but performance is compromised relative to the wild-type allele (Burch et al. 1999). This mutation may have repaired disrupted interactions between the coat and internal scaffolding proteins, especially as the mutation was a convergent change. Three different missense nucleotide substitutions and a single silent substitution were observed in gene D (table 3). The three missense substitutions are at two amino acid positions in gene D (D111 and D113). The D protein serves as an external scaffold in the morphogenesis of microvirid phages and is present in 240 copies (four for every F protein) in the phage procapsid (Hayashi et al. 1988). The external scaffolding proteins take on four unique conformations in the procapsid, labeled D1–D4 (Dokland et al. 1999). The three substitutions were within loop 5 of the D protein, which interacts with the F protein in the D1 and D2 subunits and also mediates D2–D4 and D3–D3 contacts across the 2-fold axes of symmetry (Dokland et al. 1999).

Recent work on the related microvirid phage  $\phi$ X174 suggests that rather than, or in addition to, directly improving interactions between the scaffolding proteins and the coat proteins, the compensatory mutations in genes B and D may instead be suppressing the loss of assembly intermediates into off-pathway reactions during morphogenesis. Cherwa et al. (2008) identified a number of mutations that conferred resistance to dominant lethal mutations at a particular site (residue 61) in the external scaffolding protein. These lethal mutations are dominant in that when hosts expressing the mutant proteins are infected with wild-type virus, progeny are not produced. One of the mutations

conferring resistance is at a site in the internal scaffolding gene (B103) adjacent to the mutation we observed in the ID12-ID2 lineages (B104). Uchiyama et al. (2009) identified several second site suppressors of defects in morphogenesis in response to the deletion of the first seven amino acids of the external scaffolding protein. One of these mutations was at the same site in gene D involved in recovery of the ID2-ID12Fa lineage (D111) and near a site involved in recovery of all four lineages (D113). Furthermore, several of the mutations identified by Uchiyama et al. were just upstream of the start codon for gene D near its ribosome-binding site, affecting the nucleotide sequence of gene C, similar to a mutation in the ID12-ID2Fb lineage at amino acid position 80 of gene C. Uchiyama et al. found that these mutations increased the expression of gene D. Uchiyama et al. and Cherwa et al. hypothesized that some of the mutations identified were suppressing nonproductive off-pathway reactions during morphogenesis that arise from improper interactions between the external scaffolding proteins and the coat proteins. Therefore, it may be that the primary interactions disrupted in our hybrid phages are between proteins D and F, and many of our observed compensatory mutations, including those in gene D, might be preventing the loss of viable assembly intermediates into dead-end morphogenetic pathways.

Because recovery can be achieved through mutations in just the two scaffolding proteins, the defect in the hybrids is probably manifested at some point in morphogenesis. As an extreme example of adapting to disrupted morphogenesis, Chen et al. (2007) selected the related microvirid phage  $\phi$ X174 to grow in the absence of B protein, the internal scaffold. Wild-type  $\phi$ X174 was unable to form infectious particles without the internal scaffolding protein, but through a series of substitutions in proteins A, D, F, and H as well as in promoters for gene D,  $\phi$ X174 became able to replicate without it. Therefore, redundancy is built into the system, and reduced (or absent) function of one gene can be compensated for by improved function in others. Our results are qualitatively similar. The new alleles of gene F appear to impede proper morphogenesis, but this inefficiency can be compensated for through improved function in a number of other genes involved in morphogenesis, including all the genes identified by Chen et al. (2007).

#### Compensatory evolution in the capsid and replication proteins

Four substitutions occurred in gene F, the gene that was exchanged. Interestingly, mutations at two of these sites were seen previously when the related phage, ID11, adapted to the same culturing conditions used here (Rokyta et al. 2005). The mutations yielding the amino acid substitutions F3 (V $\rightarrow$ A) in the ID12-ID2Fb lineage and F355 (P $\rightarrow$ S) in the ID2-ID12Fa lineage were at sites responding to selection in the experiments of Rokyta et al. (2005). The amino acid substitution at site F3 was different (V $\rightarrow$ F) in their experiments. These mutations may simply reflect adaptation to the culturing conditions rather than legitimate compensatory mutations, but the ancestors, ID2 and ID12, were extensively adapted to the culturing

conditions, and furthermore, these two mutations had quite large fitness effects for ID11, approximately three doublings per hour for the F3 mutation and more than five for the F355 mutation (Rokyta et al. 2005). Mutations with effects of these magnitudes would certainly have become fixed during adaptation to culture by the ancestral phages, if they had similar effects in ID12 and ID2. A mutation at F307 (N $\rightarrow$ S) in the ID2-ID12Fa lineage is at one of the sites in protein F that interacts with  $\alpha$ -helix 7 of the external scaffold, protein D, in the D4 subunit. This region constitutes the most extensive interactions between these two proteins. The final substitution in gene F was at position F217 (D $\rightarrow$ G) and is located in  $\alpha$ -helix 5 of the coat protein (McKenna et al. 1996). The only other substitution in a protein with a known structure was at position G37 (L $\rightarrow$ F) in the spike protein, encoded by gene G. The N-terminus of the spike protein is involved in the interactions between spike proteins and to a lesser extent with coat proteins (McKenna et al. 1994).

The remaining substitutions all affected sites in proteins with undetermined or undefined structures. Two different substitutions affected gene H, which encodes the pilot protein (Hayashi et al. 1988) and is a component of the mature phage virion. A single mutation-affected protein A, which is the phage replication protein and is thought to interact directly with the coat protein (Tessman and Peterson 1976; Ekechukwu et al. 1995). One of the mutations in D (D113) was also a nonsynonymous substitution in the overlapping lysis gene E. Finally, one mutation affected the coding sequence of gene C. Protein C is involved in timing of the replication cycle (Hayashi et al. 1988). Aside from being nonsynonymous in gene C, this mutation is also just upstream of the start codon for gene D and thus might be affecting the gene D ribosome-binding site as mentioned above.

Perhaps more striking than where mutations did occur is where they did not. The two proteins with the most defined and conspicuous physical interactions with the coat protein in the mature capsid played little or no role in compensatory evolution, suggesting that these interactions were not problematic in the hybrids. In the microvirid phage  $\phi$ X174, pentamers of the spike protein (G) sit at the 12 vertices formed by 60 copies of the coat protein. Each spike protein is held in place by five polar interactions and five interactions mediated by water molecules (McKenna et al. 1994; Dokland et al. 1999). The DNA-binding protein J binds to the interior of the coat proteins (McKenna et al. 1994). Nonetheless, no compensatory mutations occurred in J and only one in G. These interactions are long lived and stable, whereas the interactions with the scaffolding proteins are transient. Yet, the presumably weaker interactions are the apparent source of incompatibility between the genome and the introduced allele of gene F.

Natural patterns of genetic variation can potentially explain the lack of conflict between the exchanged alleles of the coat protein gene and the two capsid protein genes G and J. If only the G4-like group of 18 phages described by Rokyta et al. (2006) is considered, which includes both ID12 and ID2, almost no variation is evident in gene F where F interacts with proteins G and J. On the basis of the structure of the related phage  $\phi$ X174, McKenna et al.

(1994, table 5) list the sites involved in polar interactions between these proteins and the F protein. The structures of microvirid capsids are highly conserved (McKenna et al. 1996; Bernal et al. 2003), so using the structure of  $\phi$ X174 as a representative for ID12 and ID2 is not unreasonable. Of the 22 unique sites in F listed by McKenna et al. (1994) to be involved in polar interactions with either G or J, only a single site has any amino acid variation, and this site has only a single genotype with a difference out of a total of 18 genotypes within the group. These sites are conserved despite a maximum pairwise difference of  $\sim$ 12% at the amino acid level in gene F for the group. The sites in F mediating interactions with proteins J and G seem to be under strong purifying selection, and without variation at these sites, no source of conflict can arise between alleles of F and different alleles of G and J.

## Conclusions

The nature of the compensatory mutations makes clear that interactions between the coat protein and the scaffolding proteins B and D have diverged, but it is less obvious why these interactions might be different from those with G and J. First, the interactions with the scaffolding proteins are, by definition, transient. These proteins dissociate from F during maturation, so the interactions may be weaker and thus more tolerant of mutations, but as seems more germane, these regions in the coat protein are under conflicting selective pressures. The external scaffolding protein binds to the exterior of the capsid, the same region that must subsequently interact with the environment and host receptors, for example. Likewise, the internal scaffolding protein binds to the interior of the capsid, which must also later interact with protein J and the phage genome. These conflicting selective pressures may drive the divergence of the regions of interaction between the coat and the scaffolding proteins and ultimately lead to deleterious epistatic interactions in the hybrid phages. These patterns might also be found for other macromolecular complexes. Perhaps horizontal gene exchange in general is not hindered necessarily due to complex function but instead as a consequence of complex assembly.

From the structural locations of the compensatory mutations, the disrupted epistatic interactions appear to be, at least in part, manifestations of disrupted physical interactions between proteins. Interestingly, the strongest and most obvious interactions appear to be relatively unimportant sources of genic incompatibilities in hybrids because they are more likely to be evolutionarily conserved. Weaker or transient interactions, especially involving regions of genes under conflicting selective pressures, tolerate or possibly even favor variation, leading to genic incompatibilities in hybrids. This pattern is reminiscent of the well-known theoretical result for beneficial mutations suggesting that beneficial mutations of intermediate effect are most important evolutionarily (Kimura 1983; Orr 1998). Those of large effect are rare and those of small effect are often lost stochastically. Strong, stable interactions may not participate in incompatibilities because they preclude the requisite variation and are not as likely to be under conflicting selective pressures. Extremely weak

interactions do not cause incompatibilities by definition. Transient interactions, however, may allow or even favor variation and thus allow for potential genic incompatibilities in hybrids.

## Acknowledgments

The authors thank Paul Joyce, Bentley A. Fane, and Craig J. Beisel for numerous discussions. This work was supported by grants from the National Institutes of Health (P20 RR16448 and R01 GM076040).

## Literature Cited

- Bernal RA, Hafenstein S, Olson NH, Bowman VD, Chipman PR, Baker TS, Fane BA, Rossmann MG. 2003. Structural studies of bacteriophage  $\alpha$ 3 assembly. *J Mol Biol.* 325:11–24.
- Blackard JT, Cohen DE, Mayer KH. 2002. Human immunodeficiency virus superinfection and recombination: Current state of knowledge and potential clinical consequences. *Clin Infect Dis.* 34:1108–1114.
- Bull JJ, Badgett MR, Rokyta D, Molineux IJ. 2003. Experimental evolution yields hundreds of mutations in a functional viral genome. *J Mol Evol.* 57:241–248.
- Burch AD, Fane BA. 2000. Efficient complementation by chimeric Microviridae internal scaffolding proteins is a function of the COOH-terminus of the encoded protein. *Virology* 270:286–290.
- Burch AD, Ta J, Fane BA. 1999. Cross-functional analysis of the Microviridae internal scaffolding protein. *J Mol Biol.* 286:95–104.
- Charpentier C, Nora T, Tenaillon O, Clavel F, Hance AJ. 2006. Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. *J Virol.* 80:2472–2482.
- Chen M, Uchiyama A, Fane BA. 2007. Eliminating the requirement of an essential gene product in an already very small virus: scaffolding protein B-free  $\phi$ X174, B-free. *J Mol Biol.* 373:308–314.
- Cherwa JE Jr, Uchiyama A, Fane BA. 2008. Scaffolding proteins altered in the ability to perform a conformational switch confer dominant lethal assembly defects. *J Virol.* 82:5774–5780.
- Colina R, Casane D, Vasquez S, García-Aguirre L, Chunga A, Romero H, Khan B, Cristina J. 2004. Evidence of intratypic recombination in natural populations of hepatitis C virus. *J Gen Virol.* 85:31–37.
- Coyne JA, Orr HA. 2004. *Speciation*. Sunderland (MA): Sinauer Associates.
- Crill WD, Wichman HA, Bull JJ. 2000. Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* 154:27–37.
- de la Cruz F, Davies J. 2000. Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol.* 8:128–133.
- Dobzhansky T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* 21:113–135.
- Dokland T, Bernal RA, Burch A, Pletnev S, Fane BA, Rossmann MG. 1999. The role of scaffolding proteins in the assembly of the small, single-stranded DNA virus  $\phi$ X174. *J Mol Biol.* 288:595–608.
- Dokland T, McKenna R, Ilag LL, Bowman BR, Incardona NL, Fane BA, Rossmann MG. 1997. Structure of a viral procapsid with molecular scaffolding. *Nature* 389:308–313.
- Ekechukwu MC, Oberste DJ, Fane BA. 1995. Host and  $\phi$ X174 mutations affecting the morphogenesis or stabilization of the

- 50S complex, a single-stranded DNA synthesizing intermediate. *Genetics* 140:1167–1174.
- Gerrish PJ, Lenski RE. 1998. The fate of competing beneficial mutations in an asexual population. *Genetica* 102/103:127–144.
- Good JM, Handel MA, Nachman MW. 2007. Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution* 62:50–65.
- Hayashi M, Aoyama A, Richardson DL, Hayashi MN. 1988. Biology of the bacteriophage  $\phi$ X174. In: Calendar R, editor. *The bacteriophages*. Vol. 2. New York: Plenum Press. p. 1–71.
- Holmes EC, Worobey M, Rambaut A. 1999. Phylogenetic evidence for recombination in dengue virus. *Mol Biol Evol* 16:405–409.
- Kauffman SA. 1993. *The origins of order*. New York: Oxford University Press.
- Kauffman S, Levin S. 1987. Towards a general theory of adaptive walks on rugged landscapes. *J Theor Biol* 128:11–45.
- Kichler Holder K, Bull JJ. 2001. Profiles of adaptation in two similar viruses. *Genetics* 159:1393–1404.
- Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge (UK): Cambridge University Press.
- Leitner T, Escanilla D, Marquina S, Wahlberg J, Broström C, Hansson HB, Uhlén M, Albert J. 1995. Biological and molecular characterization of subtype D, G, and A/D recombinant HIV-1 transmissions in Sweden. *Virology* 209:136–146.
- Majewski J, Cohan FM. 1998. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. *Genetics* 148:13–18.
- Majewski J, Cohan FM. 1999. DNA sequence similarity requirements for interspecific recombination in *Bacillus*. *Genetics* 153:1525–1533.
- McKenna R, Bowman BR, Ilag LL, Rossmann MG, Fane BA. 1996. Atomic structure of the degraded procapsid particle of the bacteriophage G4: induced structural changes in the presence of calcium ions and functional implications. *J Mol Biol* 256:736–750.
- McKenna R, Ilag LL, Rossmann MG. 1994. Analysis of the single-stranded DNA bacteriophage  $\phi$ X174, refined at a resolution of 3.0 Å. *J Mol Biol* 237:517–543.
- McKenna R, Xia D, Willingmann P, Ilag LL, Krishnaswamy S, Rossmann MG, Olson NH, Baker TS, Incardona NL. 1992. Atomic structure of single-stranded DNA bacteriophage  $\phi$ X174 and its functional implications. *Nature* 355:137–143.
- Muller HJ. 1939. Reversibility in evolution considered from the standpoint of genetics. *Biol Rev Camb Philos Soc* 14:261–280.
- Novak CR, Fane BA. 2004. The functions of the N terminus of the  $\phi$ X174 internal scaffolding protein, a protein encoded in an overlapping reading frame in a two scaffolding protein system. *J Mol Biol* 335:383–390.
- Orr HA. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* 139:1805–1813.
- Orr HA. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52:935–949.
- Orr HA. 2006. The population genetics of adaptation on correlated fitness landscapes: the block model. *Evolution* 60:1113–1124.
- Otto SP, Feldman MW. 1997. Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theor Popul Biol* 51:134–147.
- Pepin KM, Samuel MA, Wichman HA. 2006. Variable pleiotropic effects from mutations at the same locus hamper prediction of fitness from a fitness component. *Genetics* 172:2047–2056.
- Perelson AS, Macken CA. 1995. Protein evolution on partially correlated landscapes. *Proc Natl Acad Sci USA* 92:9657–9661.
- Poon A, Chao L. 2005. The rate of compensatory mutation in the DNA bacteriophage  $\phi$ X174. *Genetics* 170:989–999.
- Poon A, Davis BH, Chao L. 2005. The coupon collector and the suppressor mutation: Estimating the number of compensatory mutations by maximum likelihood. *Genetics* 170:1323–1332.
- Poon A, Otto SP. 2000. Compensating for our load of mutations: freezing the meltdown of small populations. *Evolution* 54:1467–1479.
- R Development Core Team. 2006. *R: a language and environment for statistical computing*. Vienna (Austria): R Foundation for Statistical Computing.
- Roberts MS, Cohan FM. 1993. The effect of DNA sequence divergence on sexual isolation in *Bacillus*. *Genetics* 134:401–408.
- Robertson D, Sharp P, McCutchan F, Hahn BH. 1995. Recombination in HIV-1. *Nature* 374:124–126.
- Rokyta DR, Abdo Z, Wichman HA. Forthcoming. The genetics of adaptation for eight ssDNA bacteriophages. *J Mol Evol*.
- Rokyta D, Badgett MR, Molineux IJ, Bull JJ. 2002. Experimental genomic evolution: extensive compensation for loss of DNA ligase activity in a virus. *Mol Biol Evol* 19:230–238.
- Rokyta DR, Burch CL, Caudle SB, Wichman HA. 2006. Horizontal gene transfer and the evolution of microvirid coliphage genomes. *J Bacteriol* 188:1134–1142.
- Rokyta DR, Joyce P, Caudle SB, Wichman HA. 2005. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nature Genet* 37:441–444.
- Sabino EC, Shpaer EG, Morgado MG, Korber BTM, Diaz RS, Bongertz V, Cavalcante S, Galvao-Castro B, Mullins JI, Mayer A. 1994. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J Virol* 68:6340–6346.
- Sentandreu V, Jiménez-Hernández N, Torres-Puente M, Gosalbes MJ, Ortega E, Moya A, González-Candelas F, Bracho MA, Valero A. 2008. Evidence of recombination in inpatient populations of hepatitis C virus. *PLoS One* 3:e3239.
- Tessman ES, Peterson PK. 1976. Bacterial *rep<sup>-</sup>* mutations that block development of small DNA bacteriophages late in infection. *J Virol* 20:400–412.
- Uchiyama A, Heiman P, Fane BA. 2009. N-terminal deletions of the  $\phi$ X174 external scaffolding protein affect the timing and fidelity of assembly. *Virology* 386:303–309.
- Vulić M, Dionisio F, Taddei F, Radman M. 1997. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proc Natl Acad Sci USA* 94:9763–9767.
- Weinreich DM, Watson RA, Chao L. 2005. Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59:1165–1174.
- Yun Z, Lara C, Johansson B, de Rivera IL, Sönnnerborg A. 1996. Discrepancy of hepatitis C virus genotypes as determined by phylogenetic analysis of partial NS5 and core sequences. *J Med Virol* 49:155–160.
- Zawadzki P, Roberts MS, Cohan FM. 1995. The log-linear relationship between sexual isolation and sequence divergence in *Bacillus* transformation is robust. *Genetics* 140:917–932.

Kenneth Wolfe, Associate Editor

Accepted August 24, 2009