

Clonal Interference Is Alleviated by High Mutation Rates in Large Populations

Jonathan P. Bollback* and John P. Huelsenbeck†

*Institute of Biology, University of Copenhagen, Copenhagen, Denmark; and †Department of Integrative Biology, University of California, Berkeley

When a beneficial mutation is fixed in a population that lacks recombination, the genetic background linked to that mutation is fixed. As a result, beneficial mutations on different backgrounds experience competition, or “clonal interference,” that can cause asexual populations to evolve more slowly than their sexual counterparts. Factors such as a large population size (N) and high mutation rates (μ) increase the number of competing beneficial mutations, and hence are expected to increase the intensity of clonal interference. However, recent theory suggests that, with very large values of $N\mu$, the severity of clonal interference may instead decline. The reason is that, with large $N\mu$, genomes including both beneficial mutations are rapidly created by recurrent mutation, obviating the need for recombination. Here, we analyze data from experimentally evolved asexual populations of a bacteriophage and find that, in these nonrecombining populations with very large $N\mu$, recurrent mutation does appear to ameliorate this cost of asexuality.

Introduction

Fisher (1930) and Muller (1932) were the first to note that, in the absence of recombination, beneficial mutations initially appearing on different genetic backgrounds cannot be fixed in a population together. Because, in this scenario, only one beneficial mutation can be fixed at a time, all other beneficial mutations segregating at the same time are lost, a process called “clonal interference” (Gerrish and Lenski 1998). As a result, asexual populations suffer a reduced rate of adaptation relative to their sexual counterparts: In sexual populations (ignoring deleterious mutations), an increase in the number of mutations in a population leads to a proportional increase in the rate of adaptation. But, in asexual populations, an increase in mutation supply also increases the opportunity for clonal interference and, although the rate of adaptation increases with $N\mu$ (the product of the population size, N and mutation rate, μ), it increases more slowly than in sexual populations.

Gerrish and Lenski (1998) modeled the dynamics of beneficial mutations in a population experiencing clonal interference. Their model makes a number of predictions relevant to this study: 1) the probability of fixation of a beneficial allele declines as a function of increasing $N\mu$; 2) substitutions are “discrete, rare events, no matter how frequently beneficial mutations arise.” (Gerrish and Lenski 1998, p. 135); and 3) the expected substitution rate increases as a function of an increase in the beneficial rate, yet the increase is negligible in large populations. An important restriction of this model is that it allows a competing beneficial mutation to appear only once (Gerrish and Lenski 1998).

Several people have observed patterns of evolution consistent with clonal interference in experimental populations of microbes. Using phenotypic data, studies have shown that the rate of adaptation increases less than linearly as $N\mu$ increases (Arjan et al. 1999; Miralles et al. 1999; de Visser and Rozen 2005). Using genetic data, other studies have similarly inferred the presence of clonal interference

from dramatic reversals in the frequency of mutations, which were interpreted as beneficial mutations at first, increasing in frequency, and then being lost due to competition (Wichman et al. 1999; Holder and Bull 2001; Imhof and Schlötterer 2001). However, in the absence of haplotype data, neither kind of study can directly observe competition between beneficial mutations on different backgrounds, and the observed patterns are consistent with processes other than clonal interference (e.g., epistatic interference; Holder and Bull 2001).

However, recent theoretical work suggests that allowing for recurrent mutation can result in substantially different results. Recently, Kim and Orr (2005) modeled the dynamics of adaptation in sexuals versus asexuals at 2 beneficial sites to determine the effect of recombination on the rate of adaptation. This model differs from the Gerrish–Lenski (1998) model in that it models a discrete sequence space, which gives rise to recurrent mutation, whereas Gerrish–Lenski (1998) assume an infinite number of mutant sites, which does not allow for recurrent mutation. The differences between the 2 models are most pronounced at high $N\mu$: with very high $N\mu$, recurrent mutation can by itself bring beneficial mutations together on the same background. Hence, contrary to Gerrish and Lenski (1998), the Kim–Orr (2005) model predicts that the effect of clonal interference is greatest at intermediate $N\mu$. There are 2 particular outcomes from Kim–Orr (2005) relevant to this study: 1) the relative difference in fixation time of the second beneficial mutation, between sexuals and asexuals, is small when populations are large with a high mutation supply; and 2) as the mutation rate increases asexual populations behave progressively more like sexual populations. The Kim–Orr model suggests that substitutions in asexuals may not consist solely of discrete rare events but may occur simultaneously and more frequently.

However, both the Gerrish–Lenski (1998) and Kim–Orr (2005) models ignore the effects of deleterious mutations. For example, the predictions of the Kim–Orr model, in particular, may not apply to real organisms if an increase in μ leads to a slowdown in the rate of adaptation due to interference from deleterious mutations (Orr 2000; Wilke 2004). Here, we address this issue by studying patterns of adaptation and clonal interference in large populations (5×10^7) of a RNA virus with very high mutation rate (1.0×10^{-3} per nucleotide/replication).

Key words: adaptation, experimental evolution, clonal interference, mutation.

E-mail: bollback@binf.ku.dk.

Mol. Biol. Evol. 24(6):1397–1406. 2007

doi:10.1093/molbev/msm056

Advance Access publication March 22, 2007

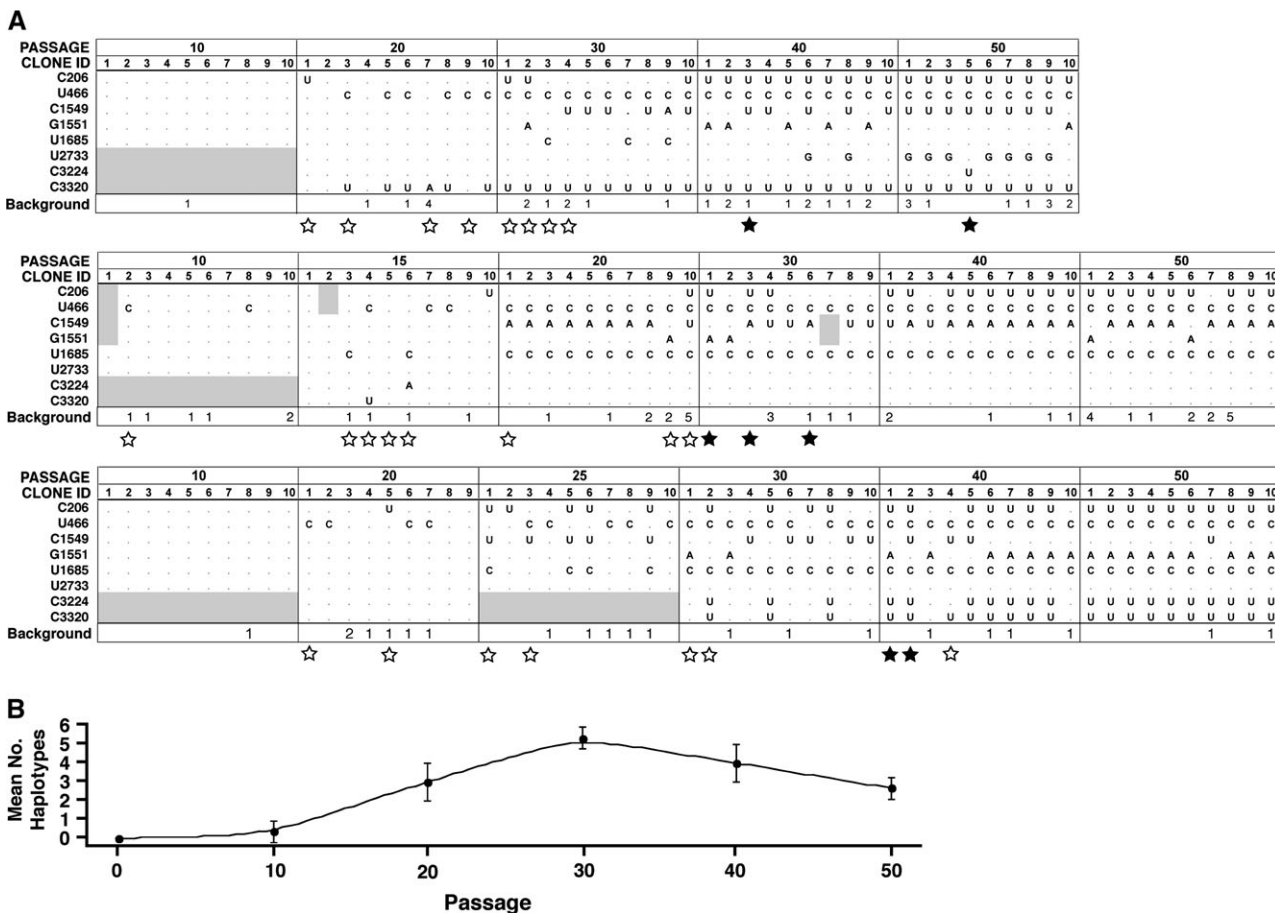


FIG. 1.—Beneficial haplotypes. (A) Haplotype structure of MS2 during adaptation. Each column represents a haplotype sampled at different time points during selection. The final row in each panel shows the number of background mutations (see supplementary material S1.polymorphisms.xls, Supplementary Material online). Unique haplotypes identified during the experiment are denoted by stars at the base of a column (open = lost, black = retained). Unknown states are shown in gray. (B) Mean number of haplotypes across all 3 evolving lines of MS2. Error bars are the standard deviation.

Methods

Phage Species

The bacteriophage species MS2 from the coliphage family Leviviridae is a positive-strand RNA bacteriophage containing 4 genes with a genome size of 3,569 nt and infects *Escherichia coli* (Murphy et al. 1995). MS2 has a mutation rate of 10^{-3} per nucleotide/replication (Drake 1993) and a recombination rate of 4.22×10^{-11} per nucleotide in multiply infected cells (Palasingam and Shaklee 1992).

Selection

The MS2 bacteriophage was selected for high temperature growth by serial passaging at 43 °C in an *E. coli* K12 host (Top10 F' strain, Invitrogen, Inc. [Carlsbad, CA]) population cultured in liquid Luria–Bertania rich medium. The host population was infected with phage at a low multiplicity of infection ($MOI \leq 0.02$ – 0.05). Three replicate lines were adapted by infecting a 3.5 ml culture of *E. coli* host cells at a density of between 2.0 – 4.0×10^8 cells per ml with 5×10^7 phages. The host population used in each passage and each line was derived from a single clonal isolate. Temperature was gradually increased during selection; phage populations were

passed at 39 °C for 10 passages of 120 min, followed by 41 °C for 10 passages of 120 min, and finally to 43 °C for 30 passages of 160 min. The additional time at 43 °C was added to account for the approximately 20% reduction in host doubling time at this temperature and to maintain a constant number of phage generations per passage.

Genome and Haplotype Sequencing

RNA was isolated from individual clones using a standard phenol–chloroform–isoamyl–ethanol extraction and amplified for sequencing using commercially available kits (SuperScript II One-Step Kit, Invitrogen, Inc.; OmniScript One-Step Kit, Qiagen, Inc. [Valencia, CA]). The RNA genomes of 5 clones from each line were sequenced using Big Dye chain termination chemistry (Applied Biosystems, Inc., Foster City, CA) to identify beneficial mutations. Clones from each line were isolated at a variety of time points and genotyped (fig. 1A).

Fitness Assays

Fitness was measured as the \log_2 of the absolute growth rate in 45 min (approximately, 1 burst) in the

selective environment (43 °C) for derived and ancestral populations. Relative fitness, w , is then defined as the ratio of the derived to the ancestral absolute growth rate as

$$w = \frac{\log_2(N_{d,1}) - \log_2(N_{d,0})}{\log_2(N_{a,1}) - \log_2(N_{a,0})},$$

where, $N_{d,i}$ and $N_{a,i}$ are the number phages sampled from populations at the i th generation for the derived (d) and ancestral (a) populations, respectively. Measures of an increase in derived population fitness were tested for significant differences using a 2-tailed Student's t -test. A generation is assumed to be 45 min for MS2 although direct measurements in our populations where not made.

Linkage Disequilibrium Test for Recombination

Evidence for recombination may be detected as a breakdown in linkage disequilibrium (LD) as a function of distance resulting in a negative relationship between these variables; recombination is more likely between physically distant sites. In addition, if recombination is free or highly frequent, we would expect to observe the absence of a correlation and, most importantly, values of the association statistic being close or equal to zero. We adopted the r^2 (Hill and Robertson 1968) measure of LD, which is defined as

$$D = f_{A_1B_1} f_{A_2B_2} - f_{A_1B_2} f_{A_2B_1},$$

$$r^2 = \frac{D^2}{f_{A_1} f_{A_2} f_{B_1} f_{B_2}},$$

where $f_{A_iB_j}$ is the frequency of mutation i at site A and mutation j at site B observed in the haplotypes, and f_{A_i} and f_{B_i} are the frequencies of mutations at sites A and B , respectively. To test for association between the LD measure, r^2 , and physical distance, we calculated Pearson's product moment correlation (r) and determined significance by performing 10,000 permutations of the distances to generate the null distribution of the correlation coefficient using the R statistics package (R Development Core Team 2005) (see supplementary material S2.permute.r, Supplementary Material online).

This analysis measured association between beneficial mutations excluding polymorphic sites as these were predominantly singletons (61 out of 66; see supplementary material S1, Supplementary Material online) and are unlikely to provide information about associations. In fact, a recent study showed little differences in estimates of the r^2 when singleton polymorphic sites were included (Piganeau and Eyre-Walker 2004).

Probability of Recombination versus Recurrent Mutation

To determine the relative contribution of recombination and mutation in generating a double beneficial mutant haplotype, we calculated the expected number of recombinants, $E[R]$, and the expected number of recurrent mutants, $E[M]$, arising in a single generation.

The following was used to calculate the expected number of recombinants in a single generation:

$$E[R] = \sum_{k=1}^{\infty} r \cdot d_{AB} \cdot \frac{\lambda^k \cdot e^{-\lambda}}{k!} \cdot \left(1 - \left(\frac{N_A + N_C}{N_T} \right)^k - \left(\frac{N_B + N_C}{N_T} \right)^k + \left(\frac{N_C}{N_T} \right)^k \right) \cdot n_{ec}$$

where k is the number of coinfecting phages; r , is the per nucleotide recombination rate (4.22×10^{-11} ; Palasingam and Shaklee 1992); d_{AB} , is the nucleotide distance between mutations; N_T , is the total number of phages; N_A , is the number of A phage haplotypes; N_B , is the number of B phage haplotypes; N_C , is the number of all other haplotypes, $N_C = N_T - (N_A + N_B)$; and n_{ec} , is the size of the host population. We calculated $E[R]$ for all possible distances between mutations in a genome the size of MS2 (see fig. 2) using the first 100 terms, $k = 2 \rightarrow 100$. The Poisson probability of observing a cell with more than 100 infecting phages is essentially 0; at an MOI of 4 the 100th term of the summation above is less than 10^{-97} .

The number of double mutants from mutation alone was calculated. Given the population size of one of the parent haplotypes, N_A , we calculate the number of double mutants as

$$E[M_A] = (1 - \mu)^{(L-2)} \cdot \mu \cdot N_A,$$

where μ , is the point mutation rate per nucleotide per replication (10^{-3} ; Drake 1993); L , is the genome length; and $(1 - \mu)^{(L-2)}$, is the probability of observing no other mutations. This calculation, however, represents only the number of double mutants arising in a single generation on the A haplotype background. Because a double mutant can also appear on the B haplotype background, the total number of double mutants, $E[M]$, is therefore

$$E[M] = (1 - \mu)^{(L-2)} \cdot \mu \cdot N_B + E[M_A].$$

The high mutation rate of MS2 gives rise to a large expected number of mutations per replication (3–4), many of which are likely to be deleterious or neutral. To account for this, we calculate the expected number of double mutants on clean backgrounds disregarding all others. The number of genomes experiencing additional mutations (or no mutations) can be seen in a comparison of the difference between total number of phage ($N_A + N_B$) with the total number of double mutants, $E[M]$, which is an exceptionally small fraction of the phage population ($2.8e^{-7}$ to $2.8e^{-5}$, see tables 1–3). For this reason, the contribution of new double mutants appearing de novo, in the fraction of the phage population (N_C) not having the beneficial mutations A or B , is ignored as this would be small (on the order of μ^2). Hence, the results presented are slightly conservative as this will only contribute to $E[M]$ and not to $E[R]$. Plots of the results (fig. 2) are displayed as the $\log_{10}(E[M]/E[R])$. An R statistical package function for performing these calculations can be found in the Supplementary Material online (S3.recombmutcompar.r).

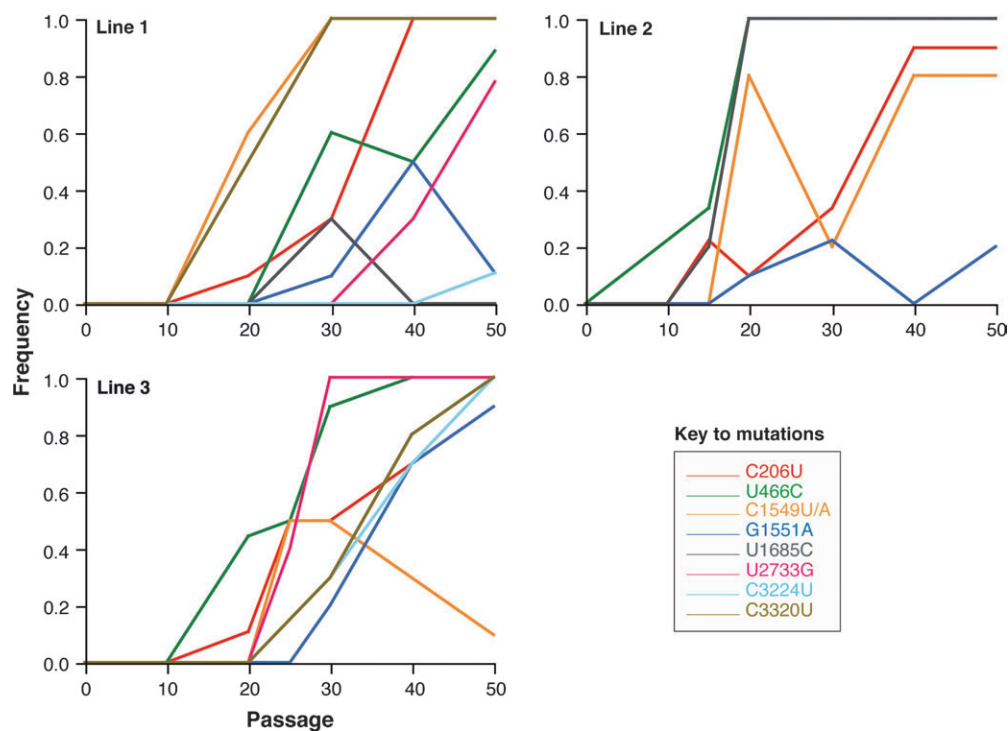


FIG. 2.—Beneficial mutation trajectories for MS2 derived lines. Mutations are identified by color as indicated in the key. Lines 1 and 3 have a C → U substitution at site 1549, whereas line 2 has a C → A substitution; in line 1 the C1549A appears in passage 30 and is lost and in line 2 C1549U is observed from passage 20–40 and then is lost. In both cases only the fixed mutation for that line is shown (see fig. 1A).

Results and Discussion

To address some of the predictions from these 2 theories, we use our data from a study in which we experimentally evolved 3 bacteriophage populations to a novel environment and tracked nucleotide substitutions occurring during adaptation. Briefly, we selected large populations ($N = 5.0 \times 10^7$) of the bacteriophage MS2, which have a high mutation rate (10^{-3} per nucleotide/replication; Drake 1993), to high temperature (43°C) growth in *E. coli* (see Methods). By sequencing full genomes at the end of selection, we identified a total of 8 fixed mutations (2 silent and 6 replacement). Each of the 8 mutations was tracked through time by sequencing regions harboring the mutations at varying time points during adaptation (fig. 1A). All the 3 populations showed dramatic increases in fitness relative to the ancestral population (at least 4.5-fold; line 1: $t_{0.05,6} = -23.665$, $P < 0.0001$; line 2: $t_{0.05,6} = -20.573$, $P < 0.0001$; line 3: $t_{0.05,6} = -17.332$, $P < 0.0001$).

Several lines of evidence suggest that the fixed mutations are beneficial as they were often parallel between lines (7/8) and included a high proportion of replacement versus silent changes (≥ 2). Furthermore, they increased rapidly in frequency within lines in a manner inconsistent with genetic drift. However, it is possible that some of these mutations instead increased rapidly in frequency via hitchhiking with beneficial mutations (Maynard Smith and Haigh 1974). The data allow us to distinguish between these possibilities as hitchhiking mutations should show complete association with at least one other mutation. Inspection of the haplotype data suggests hitchhiking is not a major factor in the evo-

lution of these populations as most of the fixed mutations (7/8) occur on multiple genetic backgrounds (see fig. 1A and the supplementary material S1.polymorphisms.xls, Supplementary Material online). In particular, the 2 silent substitutions, which would be good candidate hitchhiking mutations, also appear beneficial as they are parallel between lines and appear on different genetic backgrounds. Although the beneficial effects of the 2 silent mutations remain unclear, numerous lines of evidence suggest that silent mutations in these bacteriophages often play important roles in RNA secondary structure (van Himbergen et al. 1993) involved in gene regulation (Schmidt et al. 1987; Groenvelde et al. 1995; Kloviš et al. 1997) and encapsidation of the phage genome (Valegård et al. 1994).

By tracking haplotype changes through time, we are able to detect patterns consistent with clonal interference in all of the evolving MS2 lines (fig. 1A). Some mutations quickly reach substantial frequencies but then show dramatic reversals in mutation trajectories and extinction, apparently due to displacement by a mutation on a different genetic background (e.g., U1685C in line 1; fig. 2). In addition, many beneficial haplotypes are lost in these populations after showing increases in frequency (fig. 1A). This pattern is illustrated by the early increase in the number of beneficial haplotypes followed by a decline (fig. 1B), resulting in all but a single dominant haplotype at the end of the experiment (fig. 1A). However, negative epistasis between beneficial mutations could also explain this observation (Holder and Bull 2001). Negative epistasis occurs when 2 mutations are independently beneficial, but deleterious when combined on the same background, and has

been shown to be prevalent in RNA viruses (Sanjuán et al. 2005; Sanjuán and Elena 2006). These mutations, like those experiencing clonal interference, will compete for fixation and be displaced by mutations on different genetic backgrounds. However, other patterns in the data are inconsistent with negative epistasis causing the observed competition between the mutations. With the exception of 1549 and 1551, all of the competing mutations are observed together on the same genetic background at some point in the experiment (e.g., see passage 50 in line 3; fig. 1A). Other evidence suggests that 1549 and 1551 are in fact negatively epistatic (they cause similar and directly adjacent replacement changes in the major coat protein's FG loop) and so these sites were excluded from further analysis.

These populations, as a result of clonal interference, do suffer the loss of 18% of observed beneficial mutations (4 of 22) that have started to increase in frequency and undergo reversals in trajectory. This likely represents an underestimate of the intensity of clonal interference as some mutations may have gone unobserved and some of the polymorphic mutations (5/66) that appear in multiple lines may be beneficial (see supplementary material S1.polymorphisms.xls, Supplementary Material online). If we include the 5 polymorphic mutations that appear in more than one line, we arrive at an estimate of 33%. More importantly, during adaptation 76.7% (23 of 30) of the unique haplotypes are lost (fig. 1A).

But in spite of the high cost of clonal interference, we also see evidence that recombination or recurrent mutation is alleviating some of this cost. Contrary to the prediction of Gerrish and Lenski (1998), we do not observe sequential fixation of beneficial mutations associated with a decline in population variation during the sweep, but instead we see simultaneous sweeps of nearly all of the mutations (fig. 2). However, we do observe a buildup in haplotype variation early followed by its decline (fig. 1B). Inspection of the change in MS2 haplotype frequencies through time reveals that many of the beneficial mutations that are lost later reoccur on the dominant haplotype in the populations (fig. 1A). We address each of the possibilities—recombination or recurrent mutation—in turn.

Recombination is a simple explanation for this observation and must be seriously considered. First, the recombination rate in Leviviruses (as calculated from measurements by Palasingam and Shaklee [1992], in the related phage Q β) is extremely low, 8 orders lower than the point mutation rate (10^{-11} per nucleotide in a coinfecting cell vs. 10^{-3} per nucleotide/replication; Drake 1993). Second, recombination requires the infection of a single host by multiple phage, although this experiment was conducted with a final MOI less than 1, providing little opportunity for recombination. However, even rare recombinants can have a large effect on our observations. To address this possibility, we adopted 2 approaches: an analysis of the breakdown in LD and calculations of the relative probabilities of recombination and mutation.

First, with any appreciable level of recombination, associations between mutations will decline as a function of their proximity: physically close mutations have a lower probability of recombining than the distant ones. Analysis

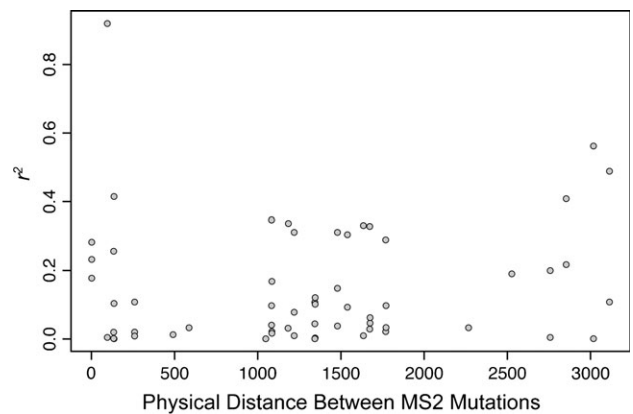


FIG. 3.—Distribution of the LD statistic, r^2 (Hill and Robertson 1968), as a function of the physical distance between mutations in MS2.

of haplotype data from all the 3 lines did not show a significant negative relationship between the LD association statistic, r^2 (Hill and Robertson 1968), and physical distance ($r = 0.0924$, $P = 0.2369$; see Methods for details). In fact, the correlation was, if anything, positive though not significantly so. In any case, the data is clearly inconsistent with high levels of recombination (fig. 3).

Second, we calculate the contribution of recombination and recurrent mutation producing a double mutant haplotype, under a variety of plausible experimental conditions (see Methods and fig. 4), to determine the relative impact of each mechanism. It is important to note that both mutations do not need to occur in the same genome replication, which has a very low probability on the order of μ^2 (10^{-6}). Instead, one of the mutations may occur at appreciable frequencies before the second mutation occurs on its background. Because these genotypes may also experience other, likely deleterious, mutations, we consider only mutations on the ancestral genetic background. The calculations show (tables 1–3 and fig. 4) that even under the most permissive conditions for recombination—high multiplicities of infection, equal frequency of the parent haplotypes, and the maximum physical distance between mutations (see fig. 4L and table 3)—recombination is expected to contribute many fewer double mutant genotypes (15,760) to the evolving population than recurrent mutation (15,786,452). When we consider the conditions that were experienced during the experiment (MOI ranging from 0.03–1.0), the relative contribution of recombination becomes negligible (using the same parameters as above but with lower MOI; tables 1–2 and figs. 4A–H), particularly for nearby sites.

In order to fully interpret the results of our calculations and their impact on our conclusions in this study, a number of issues need to be mentioned. First, as previously mentioned, we report only the expected number of beneficial double mutants appearing on clean genetic backgrounds excluding genomes with 3 or more mutations. Although high mutation rates will lead to an increased mutational load—depending on the frequency of deleterious mutations—once a beneficial mutation reaches even moderately small frequencies ($>0.055\%$), beneficial

Table 1
Relative Expected Number of Recombinants, E[R], and Recurrent Mutants, E[M], under a Low MOI Scenario

MOI	n_{cc}	N_T	N_A	N_B	E[M]	E[R]	$\log_{10}(E[M]/E[R])$
0.036	1.4×10^9	5×10^7	25000	25000	1.410	6.722×10^{-8}	7.322
0.036	1.4×10^9	5×10^7	25000	2500	0.7752	6.722×10^{-9}	8.062
0.036	1.4×10^9	5×10^7	25000	250	0.7118	6.722×10^{-10}	9.025
0.036	1.4×10^9	5×10^7	25000	25	0.7055	6.722×10^{-11}	10.02
0.036	1.4×10^9	5×10^7	25000	2.5	0.7048	6.722×10^{-12}	11.02
0.036	1.4×10^9	5×10^7	250000	250000	14.10	6.72×10^{-6}	6.322
0.036	1.4×10^9	5×10^7	250000	25000	7.752	6.721×10^{-7}	7.062
0.036	1.4×10^9	5×10^7	250000	2500	7.118	6.721×10^{-8}	8.025
0.036	1.4×10^9	5×10^7	250000	250	7.055	6.721×10^{-9}	9.021
0.036	1.4×10^9	5×10^7	250000	25	7.048	6.721×10^{-10}	10.02
0.036	1.4×10^9	5×10^7	250000	2.5	7.048	6.721×10^{-11}	11.02
0.036	1.4×10^9	5×10^7	2500000	2500000	141.0	0.000671	5.322
0.036	1.4×10^9	5×10^7	2500000	250000	77.52	6.715×10^{-5}	6.062
0.036	1.4×10^9	5×10^7	2500000	25000	71.18	6.716×10^{-6}	7.025
0.036	1.4×10^9	5×10^7	2500000	2500	70.55	6.716×10^{-7}	8.021
0.036	1.4×10^9	5×10^7	2500000	250	70.48	6.716×10^{-8}	9.021
0.036	1.4×10^9	5×10^7	2500000	25	70.48	6.716×10^{-9}	10.02
0.036	1.4×10^9	5×10^7	2500000	2.5	70.48	6.716×10^{-10}	11.02
0.036	1.4×10^9	5×10^7	2.5×10^7	2.5×10^7	1410	0.06603	4.329
0.036	1.4×10^9	5×10^7	2.5×10^7	2500000	775.2	0.006656	5.066
0.036	1.4×10^9	5×10^7	2.5×10^7	250000	711.8	0.0006662	6.029
0.036	1.4×10^9	5×10^7	2.5×10^7	25000	705.5	6.662×10^{-5}	7.025
0.036	1.4×10^9	5×10^7	2.5×10^7	2500	704.8	6.662×10^{-6}	8.024
0.036	1.4×10^9	5×10^7	2.5×10^7	250	704.8	6.662×10^{-7}	9.024
0.036	1.4×10^9	5×10^7	2.5×10^7	25	704.8	6.662×10^{-8}	10.02
0.036	1.4×10^9	5×10^7	2.5×10^7	2.5	704.8	6.662×10^{-9}	11.02

NOTE.—The MOI reflects the actual starting conditions in the MS2 selection lines. Results are shown for the maximum distance, d_{AB} , between beneficial mutations of 3,568 nt. See figure 4 for a plot of the results for all distances. See the Methods for a description of the column headers.

mutations will cooccur in a single generation on unloaded backgrounds.

Second, many of the parameter values, in particular when the frequency of the parents is small, result in an expected number of recombinants of less than one (fig. 4A–F; see also tables 1–2), whereas the expected number of double mutants arising from mutation is greater than one; mutation produces less than one double mutant only when the initial frequency of the first segregating mutant, N_A , is less than 25,000 (table 1 and fig. 4A).

Third, in our calculations of the number double-mutant genotypes generated, by recombination and recurrent mutation, we have assumed that both parental haplotypes reach the given frequencies without experiencing mutation or recombination. This assumption ignores that during the early phase of adaptation, when beneficial mutations are at low frequencies, recombination fails to generate the double mutant, yet they are being generated by mutation (tables 1–2); the double mutants are created by mutation at all but the lowest frequencies ($N_A + N_B < 0.055\%$ of the phage population), whereas generation the double mutant by recombination only begins to occur once the parental haplotypes have reached substantially higher frequencies ($N_A + N_B > 2\%$) and the population MOI is equal to or greater than unity. This latter point is important to our conclusions as our experiments rarely experienced an MOI of unity and never exceeded unity.

Finally, we must address the potential effects of the inherent uncertainty of the mutation rate estimate for MS2: the published mutation rate of 1.0×10^{-3} per nucleotide per replication (a genomic rate, U , of 3.57 per repli-

cation) for MS2 is likely to be smaller due to a fairly large confidence interval (Drake et al. 1998). In addition, most other viruses have lower mutation rates and the generality of the result needs to be considered. However, our conclusions do not change in any substantial way when mutation rates of 1.0×10^{-4} ($U = 0.35$) or 1.0×10^{-5} ($U = 0.035$) are used. In fact, a mutation rate 1 order smaller generates more double mutants on clean genetic backgrounds, whereas 2 orders lower produces fewer. This occurs because at a high mutation rate of 1.0×10^{-3} many backgrounds have other mutations, which are not counted, whereas a mutation rate of 1.0×10^{-4} results in a larger fraction of the double mutants occurring on clean backgrounds, and a mutation rate of 1.0×10^{-5} produces very few double mutants, regardless of the genetic background. (Data for these values are not shown, but the results can be obtained by adjusting E[M], in tables 1–3, by a factor of 2.48 and 0.34, respectively.) This pattern is consistent with the notion that extremely high mutation rates will lead to a decline in the rate of adaptation from interference of deleterious background mutations (Orr 2000; Wilke 2004).

Hence, although occasional recombination may have occurred in our system, our calculations suggest that recurrent mutation is both sufficient and more likely to explain the appearance of multiple beneficial mutations on a single genetic background in the different lines. Consistent with the idea that recurrent mutation occurs in these populations, 20 of the 25 fixed mutations appear on more than one background and 3 of the 66 transient polymorphic sites appear and are lost up to 3 times (see supplementary material

Table 2.
Relative Expected Number of Recombinants, E[R], and Recurrent Mutants, E[M], under an Intermediate MOI Scenario

MOI	n_{ec}	N_T	N_A	N_B	E[M]	E[R]	$\log_{10}(E[M]/E[R])$
1	1.4×10^{10}	1.4×10^{10}	7×10^6	7×10^6	394.7	0.0005267	5.875
1	1.4×10^{10}	1.4×10^{10}	7×10^6	7×10^5	217.1	5.268×10^{-5}	6.615
1	1.4×10^{10}	1.4×10^{10}	7×10^6	70000	199.3	5.269×10^{-6}	7.578
1	1.4×10^{10}	1.4×10^{10}	7×10^6	7000	197.5	5.269×10^{-7}	8.574
1	1.4×10^{10}	1.4×10^{10}	7×10^6	700	197.4	5.269×10^{-8}	9.574
1	1.4×10^{10}	1.4×10^{10}	7×10^6	70	197.3	5.268×10^{-9}	10.57
1	1.4×10^{10}	1.4×10^{10}	7×10^6	7	197.3	5.268×10^{-10}	11.57
1	1.4×10^{10}	1.4×10^{10}	7×10^7	7×10^7	3947	0.05244	4.877
1	1.4×10^{10}	1.4×10^{10}	7×10^7	7×10^6	2171	0.005255	5.616
1	1.4×10^{10}	1.4×10^{10}	7×10^7	7×10^5	1993	0.0005257	6.579
1	1.4×10^{10}	1.4×10^{10}	7×10^7	70000	1975	5.257×10^{-5}	7.575
1	1.4×10^{10}	1.4×10^{10}	7×10^7	7000	1974	5.257×10^{-6}	8.575
1	1.4×10^{10}	1.4×10^{10}	7×10^7	700	1973	5.257×10^{-7}	9.574
1	1.4×10^{10}	1.4×10^{10}	7×10^7	70	1973	5.257×10^{-8}	10.57
1	1.4×10^{10}	1.4×10^{10}	7×10^7	7	1973	5.257×10^{-9}	11.57
1	1.4×10^{10}	1.4×10^{10}	7×10^8	7×10^8	39466	5.014	3.896
1	1.4×10^{10}	1.4×10^{10}	7×10^8	7×10^7	21706	0.5128	4.627
1	1.4×10^{10}	1.4×10^{10}	7×10^8	7×10^6	19930	0.05139	5.589
1	1.4×10^{10}	1.4×10^{10}	7×10^8	7×10^5	19753	0.00514	6.585
1	1.4×10^{10}	1.4×10^{10}	7×10^8	70000	19735	0.000514	7.584
1	1.4×10^{10}	1.4×10^{10}	7×10^8	7000	19733	5.14×10^{-5}	8.584
1	1.4×10^{10}	1.4×10^{10}	7×10^8	700	19733	5.14×10^{-6}	9.584
1	1.4×10^{10}	1.4×10^{10}	7×10^8	70	19733	5.14×10^{-7}	10.58
1	1.4×10^{10}	1.4×10^{10}	7×10^8	7	19733	5.14×10^{-8}	11.58
1	1.4×10^{10}	1.4×10^{10}	7×10^9	7×10^9	394661	326.4	3.083
1	1.4×10^{10}	1.4×10^{10}	7×10^9	7×10^8	217064	40.45	3.73
1	1.4×10^{10}	1.4×10^{10}	7×10^9	7×10^7	199304	4.137	4.683
1	1.4×10^{10}	1.4×10^{10}	7×10^9	7×10^6	197528	0.4146	5.678
1	1.4×10^{10}	1.4×10^{10}	7×10^9	7×10^5	197350	0.04147	6.678
1	1.4×10^{10}	1.4×10^{10}	7×10^9	70000	197333	0.004147	7.677
1	1.4×10^{10}	1.4×10^{10}	7×10^9	7000	197331	0.0004147	8.677
1	1.4×10^{10}	1.4×10^{10}	7×10^9	700	197331	4.147×10^{-5}	9.677
1	1.4×10^{10}	1.4×10^{10}	7×10^9	70	197331	4.147×10^{-6}	10.68
1	1.4×10^{10}	1.4×10^{10}	7×10^9	7	197331	4.147×10^{-7}	11.68

NOTE.—The MOI reflects the conditions at the end of each passage in the MS2 selection lines. Results are shown for the maximum distance, d_{AB} , between beneficial mutations of 3,568 nt. See figure 4 for a plot of the results for all distances. See the Methods for a description of the column headers.

S1.polymorphisms.xls, Supplementary Material online). In addition, we observe multiple derived nucleotide states at 3 sites (fig. 1A; line 1, C1549A/U and C3320A/U; line 2 C1549A/U), and at the polymorphic site 450 in one of the lines (see supplementary material S1.polymorphisms.xls, Supplementary Material online), an observation that can only be explained by recurrent mutation.

The high mutation rate in our bacteriophage system appears to have alleviated clonal interference by allowing a rapid buildup of beneficial mutations on a single genetic background; mutations initially appear on different backgrounds, but quickly appear together and sweep to fixation. This pattern differs dramatically from the pattern observed in studies of other bacteriophage (Holder and Bull 2001; Wichman et al. 1999) that showed mostly sequential sweeps. Interestingly, in our study, compared with Wichman et al. (1999), we observe nearly a 3-fold higher substitution rate (0.05–0.06 substitutions/generation vs. 0.02 substitutions/generation). This difference could be explained by the evolution of shorter generation times in our system when host cells are in excess, and selection favors shorter generation times. Two facts suggest that the difference cannot be completely explained by this. First, the ancestral MS2 isolate used to initiate each line was preadapted for growth at the host cell densities used during selection,

and no detectable changes in fitness were observed suggesting that the phages were previously adapted to growth at high cell densities. Second, the observed difference would require a 3-fold reduction in generation time in MS2 giving rise to a generation time of nearly 15 min. The difference in substitution rates could also be explained by increased clonal interference in the study of Wichman et al. (1999) due to a larger pool of beneficial mutations (Gerrish and Lenski 1998). However, judging by dramatic changes in frequency trajectories, only a small fraction of the total mutations observed in their study appear to have been competing at any particular time (3 of 29 trajectories), suggesting that clonal interference was in fact less intense than in our study (7 of 19 trajectories). Lastly, although the populations in both experiments were similar in size ($\geq 10^7$), they experienced very different mutation rates—MS2 has a mutation rate 3 orders of magnitude higher than the $\varphi \times 174$ virus ($\approx 7.4 \times 10^{-6}$ /replication; Raney et al. 2004)—suggesting that the relative speed up was largely due to difference in mutation rates. If mutation was the primary causal agent behind the increase in substitution rate, this is consistent with the predictions of Kim–Orr (2005) model which predicts a speed up of substitutions in large asexual populations with high mutation rates. However, many factors, undoubtedly, differed

Table 3
Relative Expected Number of Recombinants, E[R], and Recurrent Mutants, E[M], under a High MOI Scenario

MOI	n_{ec}	N_T	N_A	N_B	E[M]	E[R]	$\log_{10}(E[M]/E[R])$
4	1.4×10^{11}	5.6×10^{11}	2.8×10^8	2.8×10^8	15786	0.08415	5.273
4	1.4×10^{11}	5.6×10^{11}	2.8×10^8	2.8×10^7	8683	0.008423	6.013
4	1.4×10^{11}	5.6×10^{11}	2.8×10^8	2800000	7972	0.0008423	6.976
4	1.4×10^{11}	5.6×10^{11}	2.8×10^8	280000	7901	8.423×10^{-5}	7.972
4	1.4×10^{11}	5.6×10^{11}	2.8×10^8	28000	7894	8.423×10^{-6}	8.972
4	1.4×10^{11}	5.6×10^{11}	2.8×10^8	2800	7893	8.423×10^{-7}	9.972
4	1.4×10^{11}	5.6×10^{11}	2.8×10^8	280	7893	8.424×10^{-8}	10.97
4	1.4×10^{11}	5.6×10^{11}	2.8×10^8	28	7893	8.423×10^{-9}	11.97
4	1.4×10^{11}	5.6×10^{11}	2.8×10^8	2.8	7893	8.428×10^{-10}	12.97
4	1.4×10^{11}	5.6×10^{11}	2.8×10^9	2.8×10^9	157865	8.265	4.281
4	1.4×10^{11}	5.6×10^{11}	2.8×10^9	2.8×10^8	86825	0.834	5.017
4	1.4×10^{11}	5.6×10^{11}	2.8×10^9	2.8×10^7	79722	0.08347	5.98
4	1.4×10^{11}	5.6×10^{11}	2.8×10^9	2800000	79011	0.008348	6.976
4	1.4×10^{11}	5.6×10^{11}	2.8×10^9	280000	78940	0.0008348	7.976
4	1.4×10^{11}	5.6×10^{11}	2.8×10^9	28000	78933	8.348×10^{-5}	8.976
4	1.4×10^{11}	5.6×10^{11}	2.8×10^9	2800	78932	8.348×10^{-6}	9.976
4	1.4×10^{11}	5.6×10^{11}	2.8×10^9	280	78932	8.348×10^{-7}	10.98
4	1.4×10^{11}	5.6×10^{11}	2.8×10^9	28	78932	8.348×10^{-8}	11.98
4	1.4×10^{11}	5.6×10^{11}	2.8×10^9	2.8	78932	8.348×10^{-9}	12.98
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{10}	2.8×10^{10}	1578645	692.6	3.358
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{10}	2.8×10^9	868255	75.66	4.06
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{10}	2.8×10^8	797216	7.635	5.019
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{10}	2.8×10^7	790112	0.7641	6.015
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{10}	2800000	789402	0.07642	7.014
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{10}	280000	789330	0.007642	8.014
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{10}	28000	789323	0.0007642	9.014
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{10}	2800	789323	7.642×10^{-5}	10.01
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{10}	280	789323	7.642×10^{-6}	11.01
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{10}	28	789323	7.642×10^{-7}	12.01
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{10}	2.8	789323	7.644×10^{-8}	13.01
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{11}	2.8×10^{11}	15786452	15760	3.001
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{11}	2.8×10^{10}	8682549	3304	3.420
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{11}	2.8×10^9	7972158	360.9	4.344
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{11}	2.8×10^8	7901119	36.42	5.336
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{11}	2.8×10^7	7894015	3.645	6.336
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{11}	2800000	7893305	0.3645	7.336
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{11}	280000	7893234	0.03645	8.336
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{11}	28000	7893227	0.003645	9.336
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{11}	2800	7893226	0.0003645	10.34
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{11}	280	7893226	3.645×10^{-5}	11.34
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{11}	28	7893226	3.645×10^{-6}	12.34
4	1.4×10^{11}	5.6×10^{11}	2.8×10^{11}	2.8	7893226	3.645×10^{-7}	13.34

NOTE.—The MS2 selection lines never reach an MOI of 4 and the table is provided for comparison. Results are shown for the maximum distance, d_{AB} , between beneficial mutations of 3,568 nt. See figure 4 for a plot of the results for all distances. See the Methods for a description of the column headers.

between these studies and we should be cautious drawing any solid conclusions.

What are the implications of our results for the Gerrish–Lenski (1998) and Kim–Orr (2005) models? The observation that asexual populations with a large $N\mu$ experience a simultaneous sweep of many beneficial mutations and a higher substitution rate than a population with a lower $N\mu$ is not easily reconciled with the predictions of the Gerrish–Lenski (1998) model largely because recurrent mutation is not considered by the authors. In other words, the slowdown in adaptation may be the greatest at intermediate population sizes and mutation rates, as suggested by the Kim–Orr (2005) model.

Although our results indicate that recurrent mutation allows simultaneous sweeps of beneficial mutations in asexual populations with large population sizes and high mutation rates, high mutation rates may increase other kinds of interference that can slow adaptation in asexual

populations. For example, competition between beneficial mutations due to negative epistasis may be exacerbated by high rates of recurrent mutation. Similarly, adaptation may be slowed in populations experiencing high rates of deleterious mutation (Orr 2000; Wilke 2004). Our results suggest that further experimental studies are needed to better understand under which conditions—such as, the number of available beneficial mutations, rate of epistasis, and genome size—recurrent mutation can act to alleviate the cost of clonal interference.

Supplementary Material

Supplementary Materials are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>)

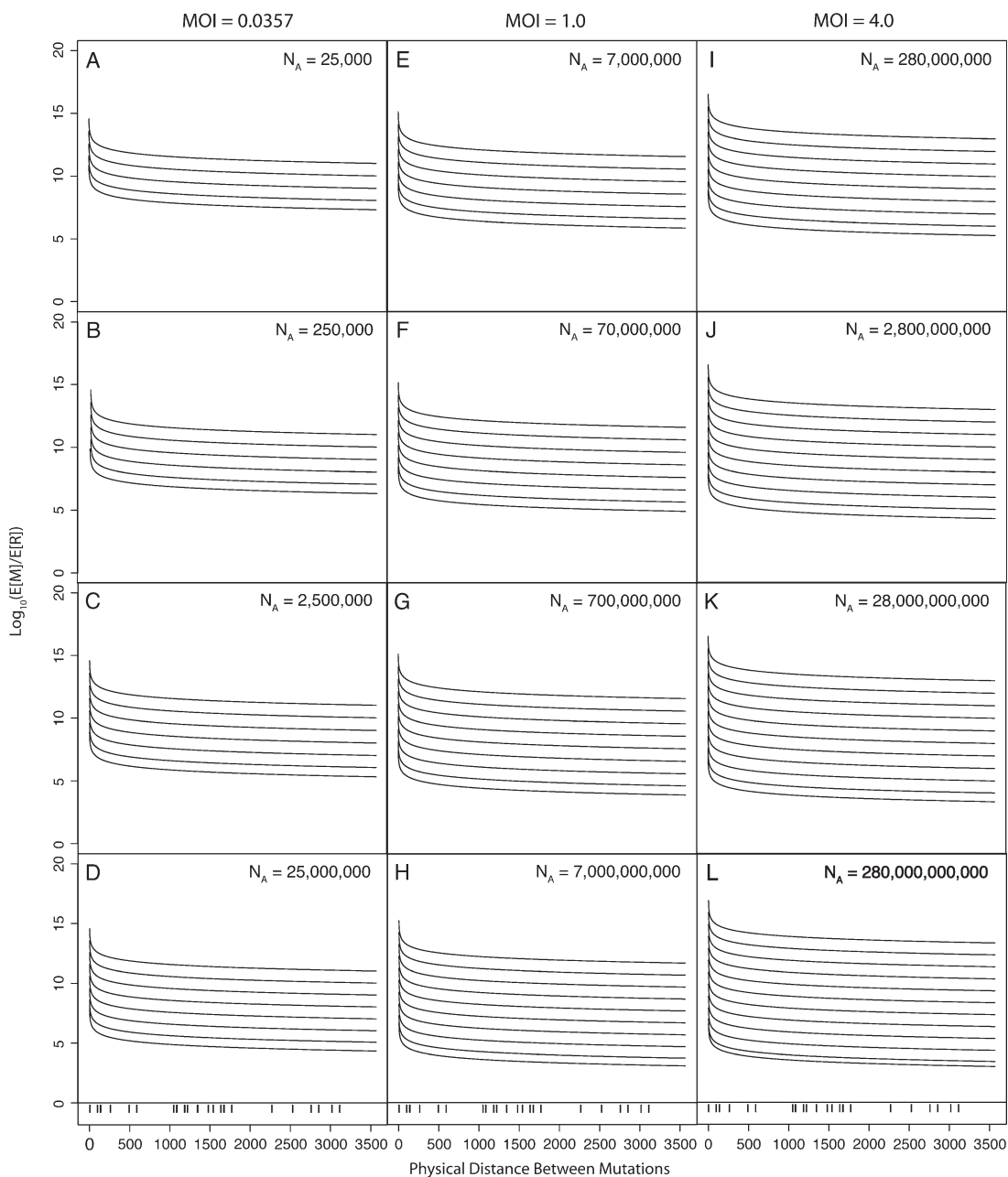


FIG. 4.—Probability of generating a double mutant, AB , arising from recurrent mutation (M) relative to recombination (R) across a variety of physical distances in the MS2 genome. Curves were calculated (see Methods) using the following parameters: N_A , the number of A genotypes; N_B , the number of B genotypes, decreasing in value from an equal frequency with N_A in 10-fold steps (the bottom line in each plot represents the condition in which A and B are at equal frequencies in the population; d_{AB} , the physical distance between mutation A and B varying from 1 to 3568, the length of the MS2 genome; $\mu = 0.001$, the rate of point mutations per replication/nucleotide (Drake 1993); $r = 4.22 \times 10^{-11}$ per coinfecting cell/nucleotide (Palasingam and Shaklee 1992); n_{EC} , the number of *Escherichia coli* host cells; N , the total number of phage; MOI, the multiplicity of infection (phage and host population sizes are increased in a manner similar to that observed in the experiment). Although the experiment never experienced an MOI > 1.0, values for an MOI of 4.0 are shown for comparison. Vertical lines at the bottom of column of plots represent the pairwise distance between all beneficial mutations observed in MS2.

Acknowledgments

Isolates of the bacteriophage were kindly provided by Dr Jan van Duin. Our thanks go to the Hoekstra Lab at the University of California, San Diego for their patience and accommodation during the sequencing phase of the project. The authors would like to thank A. Betancourt, J. Bull, D. Presgraves, H. A. Orr, Paul P. Gardner, I. Hellmann,

and R. Nielsen for helpful discussion, comments, and reviews of the manuscript. We would like to thank Barry G. Hall who provided helpful discussions and guidance during the design phase of the experiments. Our thanks also goes to 2 anonymous reviewers who provided valuable comments. This work was supported by grants from the National Science Foundation to J.P.H. (MCB-0075404) and funding from the Danish Research and Innovations Agency

(FSS Reference number 271050599) to J.P.B. during writing of the manuscript.

Literature Cited

- de Visser JAGM, Zeyl CW, Gerrish PJ, Blanchard JL, Lenski RE. 1999. Diminishing returns from mutation supply rate in asexual populations. *Science*. 283(5400):404–406.
- de Visser JAGM, Rozen DE. 2005. Limits to adaptation in asexual populations. *J Evol Biol*. 18(4):779–788.
- Drake J. 1993. Rates of spontaneous mutation rates among RNA viruses. *Proc Natl Acad Sci USA*. 90:4171–4175.
- Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. *Genetics*. 148(4):1667–1686.
- Fisher RA. 1930. *The genetical theory of natural selection*. Oxford (UK): Oxford University Press.
- Gerrish PJ, Lenski RE. 1998. The fate of competing beneficial mutations in an asexual population. *Genetica*. 102:127–144.
- Groenveld H, Thimon K, van Duin J. 1995. Translational control of maturation-protein synthesis in phage MS2: a role for the kinetics of RNA folding? *RNA*. 1:79–88.
- Hill WC, Robertson A. 1968. Linkage disequilibrium in finite populations. *Theor Appl Genet*. 38:226–231.
- Holder KK, Bull JJ. 2001. Profiles of adaptation in two similar viruses. *Genetics*. 159:1393–1404.
- Imhof M, Schlötterer C. 2001. Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. *Proc Natl Acad Sci USA*. 98(3):1113–1117.
- Kim Y, Orr HA. 2005. Adaptation in sexuals vs. asexuals: clonal interference and the Fisher-Muller model. *Genetics*. 171(3):1377–1386.
- Klovič J, Tsareva N, de Smit M, Berzins V, van Duin J. 1997. Rapid evolution of translational control mechanisms in RNA genomes. *J Mol Biol*. 265:372–384.
- Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res*. 23:23–25.
- Miralles RP, Gerrish PJ, Moya A, Elena SF. 1999. Clonal interference and the evolution of RNA viruses. *Science*. 285:1745–1747.
- Muller HJ. 1932. Some genetic aspects of sex. *Am Nat*. 66:118–138.
- Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD. 1995. *Virus taxonomy: the classification and nomenclature of viruses*. The sixth report of the international committee on taxonomy of viruses. Vienna (Austria): Springer-Verlag.
- Orr HA. 2000. The rate of adaptation in asexuals. *Genetics*. 155:961–968.
- Palasingam K, Shaklee PN. 1992. Reversion of Q β RNA phage mutants by homologous RNA recombination. *J Virol*. 66:2435–2442.
- Piganeau G, Eyre-Walker A. 2004. A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. *Heredity*. 92(4):282–288.
- R Development Core Team. 2005. *R: a language and environment for statistical computing*. Vienna (Austria): R Foundation for Statistical Computing.
- Raney JL, Delongchamp RR, Valentine CR. 2004. Spontaneous mutant frequency and mutation spectrum for gene A of phiX174 grown in *E. coli*. *Environ Mol Mutagen*. 44(2):119–127.
- Sanjuán R, Cuevas JM, Moya A, Elena SF. 2005. Epistasis and the adaptability of an RNA virus. *Genetics*. 170(3):1001–1008.
- Sanjuán R, Elena SF. 2006. Epistasis correlates to genomic complexity. *Proc Natl Acad Sci USA*. 103(39):14402–14405.
- Schmidt B, Berkhout , Overbeek GB, van Strien A, van Duin J. 1987. Determination of the RNA secondary structure that regulates lysis gene expression in bacteriophage MS2. *J Mol Biol*. 195:505–516.
- Valegård K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L. 1994. Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature*. 371(6498):623–626.
- van Himbergen J, van Geffen B, van Duin J. 1993. Translational control by a long range RNA-RNA interaction; a base substitution analysis. *Nucleic Acids Res*. 21:1713–1717.
- Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ. 1999. Different trajectories of parallel evolution during viral adaptation. *Science*. 285:422–424.
- Wilke CO. 2004. The speed of adaptation in large asexual populations. *Genetics*. 167(4):2045–2053.

Jody Hey, Associate Editor

Accepted March 16, 2007