LONG-TERM EXPERIMENTAL EVOLUTION IN *ESCHERICHIA COLI*. I. ADAPTATION AND DIVERGENCE DURING 2,000 GENERATIONS

RICHARD E. LENSKI,* MICHAEL R. ROSE, SUZANNE C. SIMPSON, AND SCOTT C. TADLER

Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92717

Submitted August 21, 1989; Revised September 11, 1990; Accepted November 2, 1990

Abstract.—We assess the degree to which adaptation to a uniform environment among independently evolving asexual populations is associated with increasing divergence of those populations. In addition, we are concerned with the pattern of adaptation itself, particularly whether the rate of increase in mean fitness tends to decline with the number of generations of selection in a constant environment. The correspondence between the rate of increase in mean fitness and the within-population genetic variance of fitness, as expected from Fisher's fundamental theorem, is also addressed. Twelve *Escherichia coli* populations were founded from a single clonal ancestor and allowed to evolve for 2,000 generations. Mean fitness increased by about 37%. However, the rate of increase in mean fitness was slower in later generations. There was no statistically significant within-population genetic variance of fitness, but there was significant between-population variance. Although the estimated genetic variation in fitness within populations was not statistically significant, it was consistent in magnitude with theoretical expectations. Similarly, the variance of mean fitness between populations was consistent with a model that incorporated stochastic variation in the timing and order of substitutions at a finite number of nonepistatic loci, coupled with substitutional delays and interference between substitutions arising from clonality. These results, taken as a whole, are consistent with theoretical expectations that do not invoke divergence due to multiple fitness peaks in a Wrightian evolutionary landscape.

The two most conspicuous features of biological evolution are adaptation of organisms to their environment and divergence of populations and species from each other. The contributions of natural selection, chance events, and historical constraints to adaptation and divergence are not easily separable. The divergence of populations may simply reflect their adaptation to different environments. Alternatively, genetic differences between populations may channel subsequent evolution so as to promote their further divergence even in the same environment (Wright 1932, 1982, 1988). And chance events, including both random genetic drift and the stochastic origin of mutations (Clarke et al. 1988), may give rise to differences between formerly identical populations subjected to uniform selection pressures (see, e.g., Cohan and Hoffmann 1986; Lenski 1988*a*).

Bacterial populations have been used for some time to study evolutionary processes, such as mutation and natural selection (see, e.g., Luria and Delbrück

* Present address: Center for Microbial Ecology, Michigan State University, East Lansing, Michigan 48824-1325.

Am. Nat. 1991. Vol. 138, pp. 1315–1341.

1943; Atwood et al. 1951). Their large population sizes and short generation times make it feasible to observe the effects of natural selection on mutations that have arisen de novo during the course of an experiment. Further, the ability to store bacteria in a nonevolving state and to maintain a strictly clonal system of propagation enables one to estimate directly the mean fitness in a particular environment of a derived population relative to that of its ancestor. Although numerous studies using bacterial populations have demonstrated increases in mean fitness due to mutation and natural selection, relatively little is known about the long-term trajectory of mean fitness, and even less is known about the dynamics of the genetic variation in fitness within and between replicate populations (Dykhuizen and Hartl 1983). Obtaining such information necessarily requires formal statistical analyses of evolutionary experiments that are both well replicated and long-term. The primary goal of our study is to provide a series of such analyses and experiments, which we hope will provide greater insight into the roles of selection, chance, and history in promoting adaptation and divergence.

In this first article, we examine experimentally the processes of adaptation and divergence in the simplest possible case: replicated populations that are founded from a common ancestral clone and then propagated in parallel in a well-defined environment. In subsequent work, we intend to consider the processes of adaptation and divergence for populations that have different evolutionary histories. In this first article, we have also restricted our analyses to a single fundamental character, the fitness of a derived clone (or mean fitness of a derived population) relative to that of its ancestor in the same environment in which the derived populations were propagated. In subsequent work, we will examine other phenotypic properties of the derived and ancestral populations, including components of fitness in that same environment, relative fitness in different environments, and metabolic characters.

Metaphorically speaking, the experiment reported here consists of starting replicate populations at the exact same point in genotypic and ecological "space" and then tracking their subsequent adaptation and divergence by following the mean and variance in the "height" of their trajectories along the adaptive surface (Wright 1932, 1982, 1988), where each population's height is measured by its mean fitness relative to that of the common ancestor. Even with such a simple and initially homogeneous experimental system, there are several plausible scenarios for adaptation and divergence. According to scenario 1, in the limit, all populations converge to the same stable adaptive peak, which may be genetically monomorphic or polymorphic. Nonetheless, stochastic variation in the time of origin of particular classes of beneficial mutants among the replicate populations gives rise to transient divergence. This first scenario predicts that mean fitness relative to the ancestral state will be strictly nondecreasing, although the rate of increase in mean fitness will decline and eventually reach zero. It also predicts that the variance in mean fitness between populations, which is initially zero, will increase for some time but will eventually drop back to zero as the rate of increase in mean fitness itself approaches zero. According to scenario 2, ecological interactions among genotypes are such that the adaptive surface itself is unstable. In particular, there may be nontransitive interactions such that, for example,

whereas clone *B* outcompetes clone *A* and clone *C* outcompetes clone *B*, clone *A* nevertheless outcompetes clone *C* (see, e.g., Paquin and Adams 1983). This second scenario predicts that mean fitness relative to the ancestral state will show some periods of decline (although not necessarily to a mean fitness below the ancestral state); the variance in mean fitness between populations may or may not eventually return to zero. According to scenario 3, replicate populations approach separate peaks of unequal mean fitness on a stable adaptive surface. The existence of the separate peaks requires strong epistasis for fitness (see, e.g., Lenski 1988*b*; Cohan et al. 1989), such that, for example, haploid genotypes *AB* and *ab* are each more fit than either *Ab* or *aB*. The divergence of the trajectories from a common starting point also depends on stochastic variation in the order of appearance of particular classes of mutants among the replicate populations. This third scenario predicts that mean fitness relative to that of the ancestor will be strictly nondecreasing, as does the first scenario. It also predicts that variance in mean fitness between populations will be sustained indefinitely, in contrast to the first scenario, wherein this variance is only transient.

In our experiment, we employ 12 replicate populations of *Escherichia coli* that were derived from a common ancestral clone. Thus, there was no initial genetic variation either within or between populations, excepting a neutral marker. These populations were allowed to evolve for 2,000 generations by serial propagation in a glucose-limited minimal medium. Mean fitness was measured at 100-generation intervals, while within- and between-population variance components were estimated every 500 generations.

MATERIALS AND METHODS

Bacterial Strains

The baseline strain used in this study is an *Escherichia coli* B clone that has also been used in several other evolutionary studies (Chao et al. 1977; Lenski and Levin 1985; Bouma and Lenski 1988; Lenski 1988*a*, 1988*b*). *Escherichia coli* B has been used in the laboratory for many decades. During this time, it has undoubtedly been cultured and stored under a variety of conditions, some of which may have been similar to those employed in the experiments reported here. Ever since this strain has been used in evolutionary studies, it has been stored frozen as a clonal isolate, essentially without evolution. We view the culture conditions in the present study as a "novel" environment, although we recognize that our founding strain may have previously adapted to similar conditions. In any case, our results seem to exclude the possibility that the founding strain was already well adapted to the culture conditions used in our experiments. This strain carries no plasmids and harbors no functional bacteriophages; it is strictly asexual. It is also prototrophic but unable to grow on the sugar L-arabinose (Ara^-). It is resistant to coliphage T6 but sensitive to other T-phages, including T5.

A spontaneous Ara^+ mutant was isolated from this strain by plating about 10^9 cells on minimal arabinose agar (Lenski 1988*a*). The Ara^- and Ara^+ clones form

red and white colonies, respectively, when they are spread on tetrazolium arabinose (TA) indicator plates (Levin et al. 1977). This colony color difference allows scoring of mixed culture composition when cultures are plated. The arabinose marker has been shown to be effectively neutral under the culture conditions used in the present series of experiments (Lenski 1988*a*), the ratio of fitnesses between the two clones being 1.00 ± 0.01 (95% confidence interval).

These two baseline clones as well as all their derivatives are stored in a glycerol-based suspension at −80°C. Hence, the baseline clones are available for direct comparison with their derivatives at any time.

Culture Conditions

The culture medium used in all experiments was Davis minimal broth (Carlton and Brown 1981) supplemented with 2×10^{-6} g thiamine hydrochloride and 0.025 g glucose per liter (Lenski 1988*a*), which permits a stationary phase bacterial density of about 5×10^7 cells per mL. Culture volume was 10 mL maintained in 50-mL Erlenmeyer flasks set in a shaking incubator at 37°C and 120 rpm. Cultures were propagated daily by transferring 0.1 mL of each culture into 9.9 mL of fresh medium. In the course of this 24-h cycle, bacterial populations attained stationary phase densities. The resulting 100-fold daily growth of each bacterial population represents ~6.64 generations of binary fission.

Fitness Assay Methods and Definitions

Relative fitness was assayed by placing two strains in competition under the liquid culture conditions described above. In all cases, one of the competitors was Ara^+ and the other was Ara^-. The two competing strains were first conditioned by growing each separately for 1 d in liquid culture. Hence, both competitors were in comparable physiological states. The two competitors were then mixed at a volumetric ratio of 1 : 1 and allowed to grow together during a standard daily growth cycle in liquid medium. Initial and final densities of each competitor were estimated by spreading them on TA plates, which permitted the competitors to be distinguished by colony color.

Let $N_1(0)$ and $N_2(0)$ be the initial densities of the Ara^- and Ara^+ competitors, respectively, and let $N_1(1)$ and $N_2(1)$ be the corresponding densities after 1 d. The average rate of increase, or Malthusian parameter, m_i, for either of the competitors during that interval is estimated as

$$m_i = \ln[N_i(1)/N_i(0)]/(1\text{ d}), \tag{1a}$$

where m_i has units of d^{-1}. A difference in two competitors' Malthusian parameters may, in principle, reflect differences in lag phase, growth rate, survival at stationary phase, or some combination of these. This estimate of the Malthusian parameter is unaffected by a strain's plating efficiency, provided that the plating efficiency is the same at $t = 0$ d and $t = 1$ d; the conditioning step described previously ensures that this provision is fulfilled. The number of doublings by either competitor, D_i, during the experimental assay is

$$D_i = \ln[N_i(1)/N_i(0)]/\ln(2), \tag{1b}$$

where D_i is a dimensionless quantity.

The fitness of one strain relative to another, W_{ij}, is estimated here as the ratio of the number of doublings of the two competitors (Lenski 1988*a*, 1988*b*; Bennett et al. 1990), which is identical to the ratio of their Malthusian parameters:

$$W_{ij} = D_i/D_j = m_i/m_j. \tag{2a}$$

We do not assign statistical significance to any single estimate of relative fitness but rather make statistical inferences based on replicated assays of fitness. The selection coefficient, S_{ij}, is given by

$$S_{ij} = W_{ij} - 1 = (m_i - m_j)/m_j, \tag{2b}$$

where W_{ij} and S_{ij} are both dimensionless quantities.

For certain purposes, we also need to estimate a selection-rate constant, r_{ij}, which reflects the difference in, rather than the ratio of, two competitors' Malthusian parameters:

$$r_{ij} = m_i - m_j = m_j S_{ij}, \tag{3a}$$

where r_{ij} has units of d^{-1} (cf. Moser 1958; Nagylaki 1977; Dykhuizen and Hartl 1983). The culture conditions used in our experiments impose strong density dependence on the overall population growth, which is limited to a 100-fold daily increase. Therefore, the average Malthusian parameter, $\overline{m}$, is

$$\overline{m} = \ln(100)/(1\ \text{d}) \cong 4.6\ \text{d}^{-1}, \tag{3b}$$

and any difference in two competitors' Malthusian parameters arises mostly, if not entirely, from "soft" selection (Wallace 1968; Christiansen 1975). This density dependence thus complicates the precise calculation of r_{ij} from S_{ij} (eq. [3a]). Nonetheless, one can reasonably approximate the average selection-rate constant as

$$r_{ij} \cong \overline{m}\, S_{ij}. \tag{3c}$$

Evolving Populations

Twelve replicate populations were founded from the two baseline clones, six Ara^- and six Ara^+. Each of the populations was founded from a single colony and, hence, a single cell; a separate colony was used for each population. Thus, over all the founding populations, there was essentially no genetic variation either within or between populations, excepting only the neutral marker.

The 12 populations were propagated for 300 d, or about 2,000 generations of binary fission, in liquid culture. Daily transfers alternated between Ara^- and Ara^+ populations, so that any cross-contamination between flasks (amounting to migration between populations) would have yielded a change in the marker state. No such changes were observed during the 2,000 generations. The initial condition of these populations, together with the lack of migration, ensured that any evolutionary changes that occurred must have resulted from changes in the frequencies of mutants arising de novo within each particular population.

Contamination Checks

The evolving populations were checked for contamination at 100-generation intervals by spreading them on TA plates. Colonies were inspected for the appro-

priate arabinose marker state. Representative colonies as well as any unusual colonies were tested for sensitivity to coliphages T5 and T6. Sensitivity to T5 demonstrates that the colony in question is indeed *E. coli,* since this phage can infect only *E. coli,* although it cannot infect all strains. Furthermore, the combination of T5 sensitivity and T6 resistance strongly indicates that the tested colonies are derivatives of the founding B strain, because it is T5 sensitive and T6 resistant, whereas most naturally occurring strains of *E. coli* are T5 resistant (R. E. Lenski, unpublished data) and most laboratory strains (e.g., K12) are T6 sensitive. In addition, both naturally occurring and laboratory strains of *E. coli* are almost invariably Ara^+, so that T5-sensitive, T6-resistant, Ara^- colonies are virtually certain to be derivatives of the founding *E. coli* B strain used in these experiments.

Contaminants were occasionally detected on the TA plates. These were always conspicuous in colony appearance as well as in possessing the red coloration of the Ara^- TA phenotype. In all cases we determined that these contaminants were present only on the plates, not in the liquid cultures, by replating from the liquid cultures and again screening for contaminants. The lack of any contaminants possessing the Ara^+ TA phenotype in the Ara^- cultures essentially excludes the possibility of such contaminants in the Ara^+ cultures. In sum, there is strong evidence against any external contamination or cross-contamination in this experiment.

Sampling

We obtained clonal and mixed population samples from the evolving populations at 100- and 500-generation intervals, respectively. Clonal samples were obtained from the same plates as those checked for contamination. For each population, a single colony was randomly chosen on the basis of its proximity to an arbitrary point on the petri dish. Each of these clonal isolates was streaked onto a slant to permit overnight growth of a dense lawn, which was then harvested and stored in a glycerol-based medium at −80°C. Mixed population samples were obtained by adding glycerol to an entire population, less that part used to propagate the culture. These were then immediately stored at −80°C.

Experimental Designs

The 100-generation clonal isolates were used to determine the trajectory of mean fitness over the entire 2,000 generations. The 500-generation mixed population samples were used to estimate the genetic variance of fitness within and between populations by using three different subsidiary designs.

Mean fitness.—The fitness of each 100-generation clonal isolate was estimated relative to the baseline strain possessing the opposite arabinose marker. The only replication was provided by the 12 populations themselves. From these 12 estimates, we computed the grand mean of fitness over all evolving populations. This design precluded partitioning of the variance components.

Variance components of fitness, design 1.—Both within- and between-population variance components were estimated by multiple clonal sampling from the stored mixed populations obtained every 500 generations. The experimental

design was hierarchical, with three levels: (1) between populations, (2) between clones, within populations, and (3) between assays, within clones. Three clones were obtained, as above, from each of the 12 populations and two fitness assays were performed on each of these clones by using the baseline clones as competitors.

Variance components of fitness, design 2.—Between-population variance of mean fitness was also estimated by using the mixed 500-generation population samples directly. This design had only two levels: (1) between populations and (2) between assays, within populations. Three fitness assays were performed by placing each of the 12 mixed-population samples in competition against the appropriate baseline clone.

Variance components of fitness, design 3.—Between-population variance of mean fitness was again estimated by using the mixed 500-generation population samples, except that in this design the competitors were the evolved populations of the opposite arabinose marker. The experimental design was such that each of the six Ara^- populations was placed in competition with each of the six Ara^+ populations, without replication. In calculating the between-population variance of mean fitness for each set of six populations, the alternative set of six was treated as a block effect. In this design, the interaction variance, if any, inflates equally the expected treatment (here, between populations) and error mean squares and, hence, biases neither the significance test nor the estimated variance component (Sokal and Rohlf 1981, pp. 349–350).

RESULTS

Mean Fitness

Temporal trajectory, linear model.—Figure 1 plots mean fitness relative to the ancestor, $\overline{W}$, against time, in generations, averaged over all 12 populations. The 95% confidence intervals are based on the t-distribution with $n - 1 = 11$ df. Mean fitness clearly increased over the 2,000-generation period. For each of the 12 populations, we performed a separate least-squares linear regression of mean fitness against time, holding the intercept at 1. The mean rate of increase in fitness relative to the ancestor was $\sim 1.9 \times 10^{-4}$ per generation over the entire 2,000 generations, with a standard deviation among the populations of only $\sim 0.2 \times 10^{-4}$ per generation (table 1). The mean fitness trajectory predicted from the linear model by using the mean rate of increase is also shown in figure 1.

Visual inspection of the actual data in figure 1 suggests that the rate of increase in mean fitness declines over the course of the experiment. To test this hypothesis, we repeated the linear regressions for each population but used only the data from the first 1,000 generations. During this interval, the mean rate of increase in fitness relative to the ancestor was $\sim 2.7 \times 10^{-4}$ per generation (table 1). We also computed, by subtraction, the mean rate of fitness increase during the last 1,000 generations for each population. Mean fitness continued to increase significantly, but at the lower rate of $\sim 1.1 \times 10^{-4}$ per generation (table 1). In all 12 populations, the rate of fitness increase was greater in the first 1,000 generations

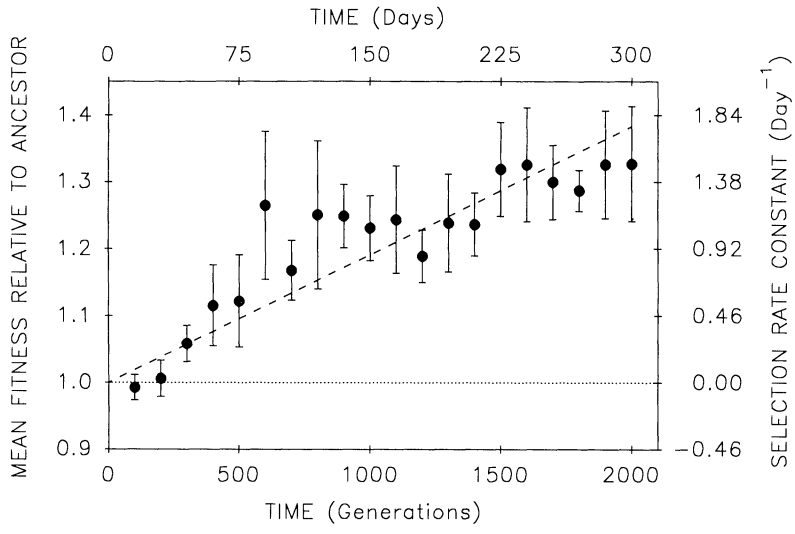

FIG. 1.—Trajectory of mean fitness during 2,000 generations (300 d). Fitness is expressed relative to an ancestral strain that has a neutral genetic marker; the selection-rate constant is related to fitness as described by eqq. (2b) and (3c). Mean fitness was estimated for each of the 12 evolving populations at 100-generation intervals. *Filled circles*, grand means of 12 estimates of mean fitness. The error bars show the 95% confidence interval based on the t distribution with 11 df. *Dashed line*, least squares linear regression of mean fitness against time, with the intercept constrained to 1. The slope of this line was calculated as the mean of the 12 slopes obtained from separate regressions for each of the 12 populations (table 1).

than in the second 1,000 generations, and the resulting difference is highly significant ($p < .001$; table 1).

Temporal trajectory, step model.—We have a priori reasons, however, for doubting the strict applicability of a linear model to the evolution of mean fitness in clonal populations. As a consequence of the linkage disequilibrium resulting from clonality, advantageous mutations must be incorporated into an evolving population sequentially, rather than simultaneously (Muller 1932). The dynamics of each selective substitution in a sequence implies an almost steplike trajectory for mean fitness. Consider the appearance and subsequent fixation of a single favorable mutation in one population. When the mutation first occurs, the frequency of the mutant genotype, $P(0)$, is equal to $1/N$, where N is the population size. (In the present experiment, the populations are serially cultured, so that N cycles from $\sim 5 \times 10^6$ to $\sim 5 \times 10^8$ and the initial frequency of a new mutant ranges correspondingly from $\sim 2 \times 10^{-9}$ to $\sim 2 \times 10^{-7}$.) Assuming that the favorable new mutation is not lost by drift (as it often will be), then the rate of change in the frequency of the allele is governed by

$$dP/dt = r_{ij}P(1 - P) \tag{4}$$

(Nagylaki 1977, p. 20), where r_{ij} is the selection-rate constant, that is, the difference in the Malthusian parameters of the favorable mutant and its progenitor (eq. [3a]). Mean fitness in the population, $\overline{W}(t)$, depends on the frequency of the

TABLE 1

ANALYSES OF TRAJECTORIES FOR MEAN FITNESS IN TWELVE POPULATIONS

	Mean ± SE	t_s
Linear model:		
Rate of change in mean fitness (per 1,000 generations):		
A. During all 2,000 generations	.192 ± .006	34.36***
B. During first 1,000 generations	.275 ± .015	18.10***
C. During last 1,000 generations[a]	.108 ± .020	5.41***
D. Difference[b]	.167 ± .034	4.95***
Step model:		
Net change in mean fitness:		
A. During all 2,000 generations	.368 ± .033	11.24***
B. During first 1,000 generations	.241 ± .009	27.12***
C. During last 1,000 generations	.126 ± .034	3.74**
D. Difference[b]	.115 ± .037	3.13**
Number of steps:		
A. During all 2,000 generations	3.667 ± .284	12.90***
B. During first 1,000 generations	2.500 ± .151	16.58***
C. During last 1,000 generations	1.167 ± .207	5.63***
D. Difference[b]	1.333 ± .225	5.93***
Average step size:		
A. During all 2,000 generations	.103 ± .008	13.77***
B. During first 1,000 generations	.101 ± .008	12.52***
C. During last 1,000 generations[c]	.103 ± .014	7.40***
D. Difference[b,c]	−.004 ± .016	−0.25 (NS)

NOTE.—The mean and SE were calculated from 12 replicate populations. The t statistic (t_s) and significance level were based on a two-tailed test for the null hypothesis that the mean is equal to zero, using the t distribution with 11 df. For both models, the intercept (i.e., initial mean fitness) was fixed at 1. NS, Not significant.

[a] Rate of change during last 1,000 generations was calculated as twice quantity A minus quantity B.

[b] Difference was calculated as quantity B minus quantity C.

[c] Excluding two populations for which no steps were observed during the last 1,000 generations, and by using the t distribution with 9 df.

** $p < .01$.

*** $p < .001$.

favored mutant according to

$$\overline{W}(t) = 1 + S_{ij}P(t) \cong 1 + r_{ij}P(t)/\overline{m}, \quad (5)$$

where S_{ij} is the selection coefficient (eq. [2*b*]) and $\overline{m}$ is the average Malthusian parameter (eq. [3b]). In figure 2, we have plotted the trajectory of mean fitness for mutants with selection coefficients of 0.05, 0.1, and 0.2. In all cases, the favorable mutation takes many generations to reach a frequency at which it has any appreciable effect on mean fitness, followed by a comparatively few generations during which there is a more rapid increase in mean fitness, with a plateau at a mean fitness equal to $1 + S_{ij}$. (Of course, both the apparent delay and the period of appreciable increase are longer for smaller S_{ij}.) Before the first substitution and each substitution thereafter, there will also be a further delay—which may be short or long—associated with waiting for the stochastic appearance of the favorable mutation (again, those many mutations lost by drift shortly after they appear are ignored). Therefore, our a priori expectation for the evolution of

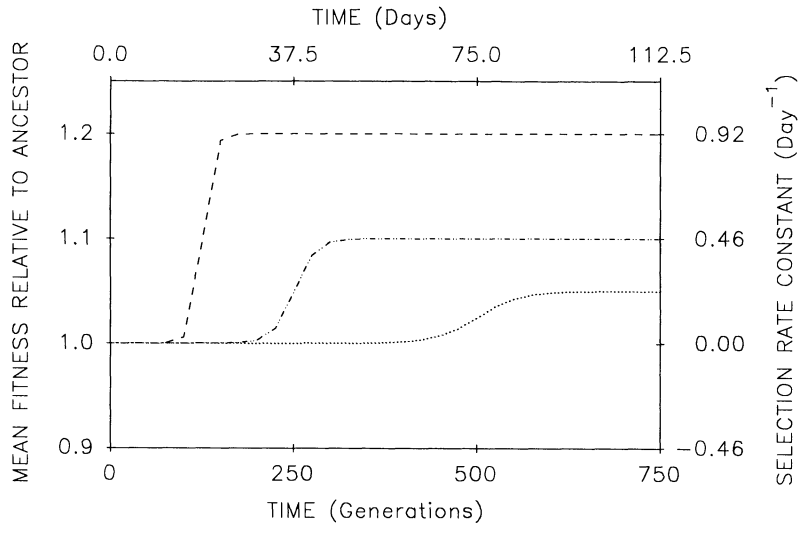

FIG. 2.—Steplike trajectory for mean fitness during the substitution of a favorable new mutation. The three lines correspond to mutants with selection coefficients, S_{ij}, equal to 0.05 (*dotted line*), 0.1 (*dots and dashes*), and 0.2 (*dashed line*). For all three lines, the initial frequency of the favored mutant, $P(0)$, was set to 3×10^{-8}. In each case, the favored mutant takes many generations to reach a frequency at which it has any appreciable effect on mean fitness, followed by a comparatively few generations during which there is a more rapid increase in mean fitness, with a plateau at mean fitness equal to $1 + S_{ij}$.

mean fitness is a series of steplike increases in each population (see also Ginzburg 1981).

We have fitted separately each of the 12 populations' fitness trajectories to a step function by using an isotonic regression model (Robertson et al. 1988; Gaines and Rice 1990) as follows. The pool-adjacent-violators algorithm (Robertson et al. 1988) was used to amalgamate consecutive fitness estimates into the largest number of groups fully consistent with the alternative hypothesis that each group mean is greater than all preceding group means. (We further constrained the intercept, or initial mean fitness, to be equal to 1.) For each population, this "full" model consisted of six to nine groups, including the initial group, corresponding to five to eight steps upward in mean fitness. Each full model could be further described by an error sum of squares (i.e., resulting from deviations of the sample fitness values from the group, or step, means), which was some fraction of the total sum of squares (i.e., resulting from deviations of the sample fitness values from the mean fitness under the null hypothesis, which equals 1). The full model is "overfitted," in that it includes all steps consistent with the hypothesis that there is a strictly nondecreasing mean fitness, regardless of whether the difference in means between adjacent groups is small or large relative to the mean square error. Therefore, we used a backward elimination procedure (Kleinbaum and Kupper 1978) to combine adjacent groups until we had identified some smaller number of steps such that each remaining step had an associated partial F statistic that was significant at $p < .5$. (The choice of significance level is necessarily somewhat arbitrary. The important point here is that we sought to balance Type

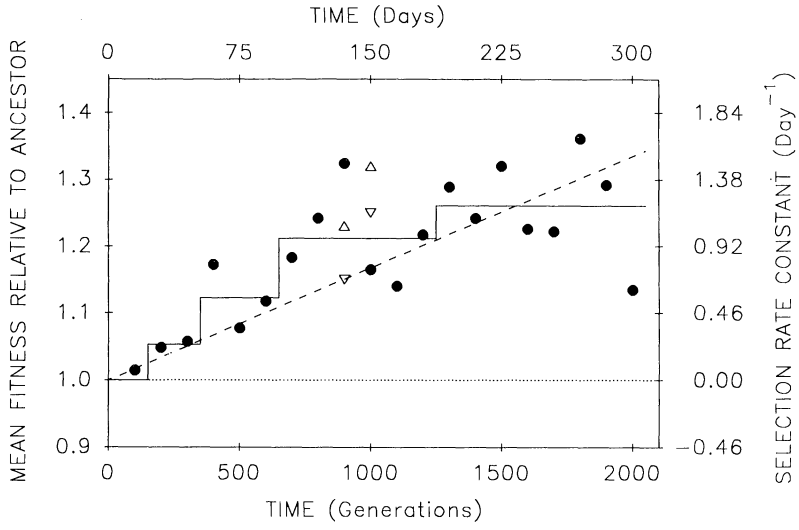

FIG. 3.—Trajectory of mean fitness for one of the 12 replicate populations. *Filled circles,* mean fitnesses, as originally estimated. *Dashed line,* least squares linear regression for the original estimates, with the initial mean fitness constrained to 1. *Solid line,* fit of the step model, obtained by isotonic regression as described in the text, again for the original estimates and with the intercept fixed at 1. The largest apparent drop in mean fitness occurred between the 900- and 1,000-generation samples. *Hollow triangles,* mean fitnesses as estimated by repeating the assays for the 900- and 1,000-generation samples. The drop in mean fitness is no longer apparent. Similar results obtained over all 12 populations support the hypothesis that observed drops in mean fitness represent experimental or sampling errors, and do not indicate nontransitive interactions in determining fitness.

I errors, in which a steplike increase in mean fitness is claimed when it did not occur, and Type II errors, in which a steplike increase in mean fitness occurred but is not claimed. A more stringent level, such as $p < .05$, clearly may exclude many real steps, while an even less stringent level could, like the full model, include some spurious steps.) For each population, the resulting condensed model had two to five step increases in mean fitness. The trajectory of mean fitness is shown in figure 3 for one of the populations, along with the corresponding step and linear models. One can compute a mean square associated with the improved fit of the step model, relative to that of the linear model, by dividing the difference in the sum of squares explained by the two models by the difference in the degrees of freedom required by each; a partial F statistic is then calculated by dividing this mean square for the marginal improvement of the step model by its mean square error. For the population shown in figure 3, the marginal improvement of the step model relative to the linear model is significant at $p < .05$. In all 12 populations, the corresponding improvement in fit is significant at $p < .1$ or less (for two of the populations, $.05 < p < .1$; for three others, $.01 < p < .05$; for six populations, $.001 < p < .01$; and for one population, $p < .001$). Evidently, the step model provides a significantly improved fit to the trajectories of mean fitness.

The step model, averaged over all 12 populations, indicates a net increase in mean fitness of ~0.37 during the 2,000 generations, with almost two-thirds of this

increase having occurred during the first 1,000 generations (table 1). By using this step model, one can also ask whether the diminution in the rate of fitness increase reflects primarily a reduction in the average step size or an increase in the waiting time between successive steps. The observed decline in the rate of fitness increase was evidently due to a decrease in the number of steps during the second 1,000-generation interval, rather than to any discernible reduction in the average step size (table 1). This result suggests that there may be yet to come additional substitutions with comparably high selection coefficients ($S \cong 0.1$), albeit at less frequent intervals. If so, then the apparent fitness plateau between generations 1,500 and 2,000 could be illusory, although the significant decline in the rate of fitness increase during the second 1,000 generations must be regarded as real.

Possible nontransitivity of fitness interactions.—The preceding analyses, including especially the isotonic regression, are predicated on the assumption that fitness interactions between successive clones are transitive. However, Paquin and Adams (1983) have demonstrated nontransitive fitness interactions among clones from evolving populations of the yeast *Saccharomyces cerevisiae*. In their study, some evolved clones were actually less fit than the baseline strains from which they were derived, although all clones were apparently more fit than their immediate predecessors. We therefore sought to examine the validity of the assumption of transitive fitness interactions among the successive *Escherichia coli* clones obtained in this study.

Mean fitness over all 12 populations never dropped appreciably below that of the ancestor (fig. 1). And although mean fitness estimates sometimes declined slightly relative to earlier estimates, none of the 190 possible paired-comparison t-tests indicates a significant decline ($p > .05$, even without adjusting for the "table-wide" error rate associated with performing so many tests [Rice 1989]).

Apparent drops in fitness were more conspicuous in the individual trajectories for the 12 replicate populations. We performed the following experiment in order to determine whether such drops were due to experimental and sampling errors, or whether such drops represented actual declines in fitness between successive clonal isolates, as measured relative to the ancestor. For each of the 12 fitness trajectories, we identified the pair of consecutive sample points that produced the largest apparent drop in fitness (e.g., the 900- and 1,000-generation samples for the population depicted in fig. 3). Each of the two corresponding clonal isolates was again placed in competition with the ancestor, according to a paired-comparisons design with twofold replication. The null hypothesis was that the apparent drops in fitness represented experimental and sampling errors and should not be reproducible. Accordingly, one would expect the earlier and later clonal isolates to split 50% of the doubly replicated paired comparisons, whereas the earlier and the later isolates should each prevail in *both* replicated comparisons with a frequency of 25%. Alternatively, if there were real declines in fitness relative to the ancestor, then the earlier clonal isolates should also prevail in the subsequently repeated fitness assays.

One set of repeated fitness assays is shown in figure 3, along with the original fitness estimates. In the repeated fitness assays, there is no indication that the earlier clonal isolate is any more fit than the later isolate. For the entire experiment involving all 12 populations, the outcome of the repeated fitness assays

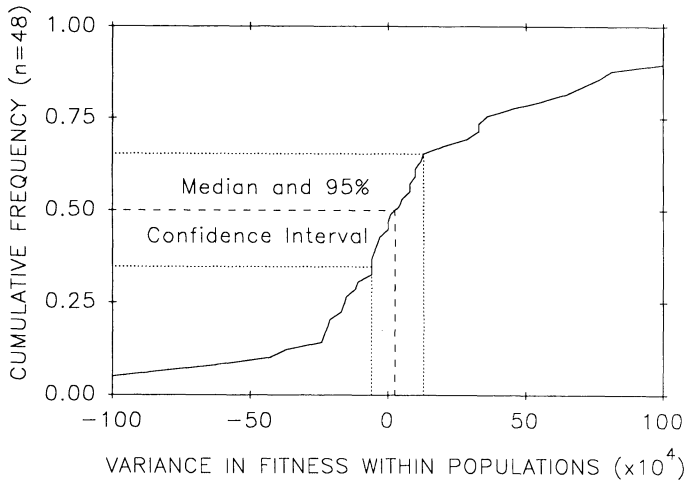

FIG. 4.—Cumulative frequency distribution of the estimated within-population genetic variances of fitness. Estimates were obtained after 500, 1,000, 1,500, and 2,000 generations for each of the 12 independently evolving populations. The distribution-free 95% confidence interval about the sample median of the 48 estimates includes zero. The confidence interval also includes the expected values of the genetic variance of fitness within populations calculated from the rate of change in mean fitness over either the entire 2,000-generation experiment ($\sim 2.7 \times 10^{-4}$) or just that period of most rapid increase in mean fitness from generations 200 to 600 ($\sim 9.4 \times 10^{-4}$).

conformed exactly to the expectation under the null hypothesis: in three cases, the earlier clonal isolate prevailed; in six cases, the earlier and later isolates split the two paired comparisons; and in three cases (including that shown in fig. 3), the later clonal isolate prevailed. In sum, we find no evidence to indicate nontransitivity of fitness interactions in this study.

Variance in Fitness within Populations

Design 1 was used to estimate the genetic variance in fitness between clones within each of the 12 populations after 500, 1,000, 1,500, and 2,000 generations, as described in the Materials and Methods section. Variances were estimated from the difference between the mean squares between clones (within populations) and between assays (within clones) divided by the number of replicate assays per clone (two).

Only three of the 48 estimates of clonal variation for fitness within populations were significant at $p < .05$ (for two of these, $.01 < p < .05$; for one, $.001 < p < .01$). There was some suggestion that an excessive number of analyses were significant at $p < .25$, but a Kolmogorov-Smirnov test (Sokal and Rohlf 1981) comparing the cumulative frequency distribution of observed and expected significance levels (i.e., p values) was itself not significant ($p > .2$). A Kruskal-Wallis test gave no indication that the ranked variances were heterogeneous across the four sample periods ($p > .5$).

The cumulative frequency distribution of the estimated variances of fitness between clones within populations is shown in figure 4. These 48 variance esti-

mates are not expected to be normally distributed; therefore, we constructed a distribution-free 95% confidence interval about the median of these variance estimates (Mosteller and Rourke 1973). The median of the 48 variance estimates is 2.5×10^{-4}, while the confidence interval runs from -6×10^{-4} to 13×10^{-4}. Our conclusion is that we have been unable to detect statistically significant levels of within-population genetic variance of fitness.

This failure to detect significant variance of fitness within populations is seemingly anomalous, given that we have detected significant increases in mean fitness. During the periods of selective substitution, the populations must have been polymorphic. Is the low level of within-population variance inconsistent with this inferred polymorphism? Consider a single selective substitution. Equation (4) gives the rate of change in the frequency of a favored allele, $P(t)$, while equation (5) describes the dependence of mean fitness, $\overline{W}(t)$, on the frequency of the favored allele. From these, one obtains the rate of change in mean fitness as

$$d\overline{W}/dt = (d\overline{W}/dP)(dP/dt) = SrP(1 - P) \cong \overline{m}S^2P(1 - P) = \overline{m}\,\mathrm{Var}(W). \quad (6)$$

The *cumulative* variance of fitness within a population between times t and t' is given by

$$\begin{aligned} \int_t^{t'} \mathrm{Var}(W)\,dt &= (1/\overline{m}) \int_t^{t'} S(dP/dt)\,dt \\ &= (1/\overline{m}) \int_{P(t)}^{P(t')} S\,dP \\ &= (1/\overline{m})\,S\,[P(t') - P(t)] \\ &= (1/\overline{m})\,[\overline{W}(t') - \overline{W}(t)]\,; \end{aligned} \quad (7a)$$

that is, the total variance of fitness during the course of a selective substitution is proportional to the change in mean fitness (see also Crow 1968; Felsenstein 1971). Consequently, the expected value of the within-population variance of fitness between times t and t', $E(\mathrm{Var}(W))$, is given by

$$\mathrm{E}(\mathrm{Var}(W)) = (1/\overline{m})\,[\overline{W}(t') - \overline{W}(t)]/(t' - t). \quad (7b)$$

If one assumes further that the selection coefficients are additive, then equation (7b) can also be applied to time intervals comprising two or more successive substitutions (including also the period between substitutions, in which the population is monomorphic). In this experiment, the average Malthusian parameter was $\sim 4.6\ \mathrm{d}^{-1}$ (eq. [3b]), the total increase in mean fitness was ~0.37 (table 1), and the time elapsed was 300 d, so that $\mathrm{E}(\mathrm{Var}(W)) = 2.7 \times 10^{-4}$. This theoretical expectation is very close to the median of the empirical estimates of $\mathrm{Var}(W)$, 2.5×10^{-4}. Hence, there is no inconsistency between the within-population variance of fitness that is expected to exist during the course of these selective substitutions and that which was observed, the small magnitude of the latter notwithstanding.

The preceding analysis might be criticized on the grounds that the expected variance is assumed to be constant in time, when clearly it need not be. We can extend this analysis by estimating the expected variance in fitness within

TABLE 2

ESTIMATES OF VARIATION IN MEAN FITNESS BETWEEN POPULATIONS ($\times 10^4$)

Generation	500	1,000	1,500	2,000
Ara$^-$ set of populations:				
Design 1	14.9	1.8	35.6	10.0
Design 2	36.5**	−12.3	31.4$^+$	−9.1
Design 3	−13.2	9.4*	5.8*	95.1**
Ara$^+$ set of populations:				
Design 1	29.3**	39.5*	−21.1	31.8**
Design 2	8.3	6.8	−1.8	9.3$^+$
Design 3	3.4	6.0$^+$	14.5***	152.4***

$^+ p < .1$.
$^* p < .05$.
$^{**} p < .01$.
$^{***} p < .001$.

populations at particular times by using the theoretical correspondence between the rate of change in mean fitness and the variance of fitness within populations, as given by equation (6). Consider the period of most rapid change in mean fitness, between generations 200 and 600 (days 30 and 90; fig. 1). The ~26% increase in mean fitness during these 60 d gives an expected within-population variance of 9.4×10^{-4}, which is still well within the confidence interval. In conclusion, the observed within-population genetic variance of fitness is wholly consistent with theoretical expectations, given the observed trajectory of mean fitness.

Variance of Mean Fitness between Populations

Designs 1, 2, and 3 were used to estimate the variance of mean fitness between populations after 500, 1,000, 1,500, and 2,000 generations, as described in the Materials and Methods section. Analyses were performed separately for the six Ara$^-$ and the six Ara$^+$ populations in order to provide two independent estimates of the between-population variance for each design at each sample time. (Also, for design 3, mean fitnesses for the two groups were obtained relative to different sets of competitors.)

The estimates of between-population variance are given in table 2 for each of the arabinose marker groups, each of the generations assayed, and each of the designs. While there is a great deal of heterogeneity in the estimated variances, two conclusions can be drawn from these results. First, there is an excess of statistically significant variances. Nine of the 24 values are significant at $p < .05$. Moreover, a Kolmogorov-Smirnov test comparing the cumulative frequency distribution of observed and expected significance levels indicates a highly significant ($p \ll .01$) excess of low p values. Second, there is no clear pattern of differentiation between the two sets of populations, the three designs, or across times. The median of the 24 estimates for the between-population variance is 9.4×10^{-4}, while the distribution-free 95% confidence interval for the median (Mosteller and Rourke 1973) runs from 3.4×10^{-4} to 31.4×10^{-4}.

Unfortunately, it is also not clear whether this variation in mean fitness between populations is transient or will be sustained. Again, there is the question of what are the theoretical expectations. Specifically, what is the expected time course of the between-population variance of mean fitness during a series of selective substitutions? We will consider three alternative models, all of which assume that selection coefficients are additive (i.e., nonepistatic).

Model A, simple Poisson process.—Consider a population cultured serially, as in these experiments, with initial population size N_0 ($\sim 5 \times 10^6$) and final population size N_f ($\sim 5 \times 10^8$) over the culture cycle. The number of generations of binary fission during a single cycle of growth, g, is $\log_2 (N_f/N_0)$, which equals ~ 6.64 in this experiment. The arithmetic mean population size over the course of this cycle, $\overline{N}$, is approximated by $2(N_f - N_0)/g$, which equals $\sim 1.5 \times 10^8$. Now assume a constant rate of favorable mutation per cell generation, u, each mutation having an equal selection coefficient, S. The rate of favorable mutations per generation, M, including those lost to drift, averaged over the entire growth cycle, is equal to $u\overline{N}$. R. R. Hudson (personal communication) has shown that the substitution rate for favorable mutations per generation, k, can be approximated as

$$k \cong u\, N_0\, g\, 2S, \tag{8a}$$

when S is a small positive value. Hence, the average probability of eventual fixation for a favorable mutation, f, is given by

$$f = k/M \cong u\, N_0\, g\, 2S/u\overline{N} = 2S\, N_e/\overline{N}, \tag{8b}$$

where the effective population size, $N_e = N_0\, g$, is approximately equal to the harmonic mean of the population sizes during the growth cycle (see also Haldane 1927; Wright 1938; Moran 1962; Kimura 1983). Here, N_e is equal to $\sim 3.3 \times 10^7$.

If these adaptive substitutions occur randomly and independently over a period of t generations, then they should be distributed according to the Poisson with mean kt (Feller 1957). Therefore, the variance between populations of the number of substitutions after t generations should also equal kt. The expected total fitness gain during t generations is thus ktS, while the expected variance between populations of the total fitness gain is ktS^2 (J. Felsenstein, personal communication). The step model fitted to the fitness trajectories yielded an average of $kt \cong 3.67$ substitutions per population during 2,000 generations and an average selection coefficient of $S \cong 0.10$. From these values and equation (8*a*), we infer a corresponding rate for these favorable mutations, u, equal to $\sim 2.8 \times 10^{-10}$ per cell generation. We can also use these relations to compute an expected variance of mean fitness between populations, $\mathrm{E}(\mathrm{Var}(\overline{W}))$, which increases linearly from zero, at the start of the experiment, to 3.67×10^{-2}, after 2,000 generations (fig. 5). This $\mathrm{E}(\mathrm{Var}(\overline{W}))$ is much greater than the corresponding observed variance (table 2). Model A would be consistent with the observed variance only if one assumed a much larger number of advantageous substitutions per population during the 2,000 generations ($kt \cong 30$), each one providing only a small selective advantage ($S \cong 0.01$). Model A is also deficient in that it predicts a constant rate of substitution (k), and hence a constant rate of change in mean fitness (kS),

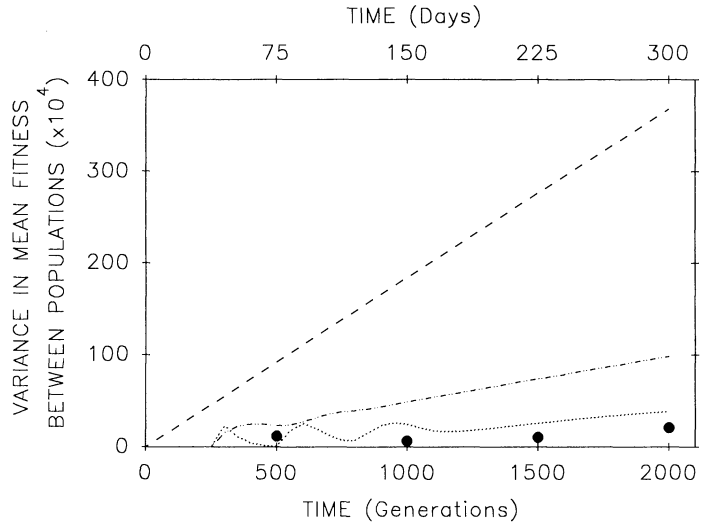

FIG. 5.—Theoretical and observed genetic variances of mean fitness between independently evolving populations. *Filled circles,* median of the six variance estimates at each sample time (table 2). The three lines give the expected variances obtained from different models of the substitution process; the parameters for all three models were chosen to produce an expectation of 3.67 substitutions of fitness-enhancing alleles during 2,000 generations (fig. 6), each with an associated selection coefficient, S, equal to 0.1. Model A (*dashed line*) represents a simple Poisson process. Model B (*dots and dashes*) incorporates a delay of 250 generations associated with the rise of each new mutant to a high frequency (fig. 2); during each such interval, other fitness-enhancing alleles are excluded from fixation because of the lack of recombination among genomes. Model C (*dotted line*) includes the same delay and interference among clones as in model B; however, in model C, the number of distinct classes of fitness-enhancing alleles is finite, and the mutation rates for these classes can vary. Further details for each model are provided in the text.

whereas the observed pattern of substitution (fig. 6) and trajectory of mean fitness (fig. 1) show an initial delay and then a subsequent diminution in rate.

Model B, Poisson process with delays.—Model A ignores the delay associated with the rise of a new mutation from some low initial frequency to a frequency at which the mutant appreciably affects mean fitness (fig. 2), and it implicitly assumes that selective substitutions occur independently of one other. However, in the case of selection acting on asexual populations, multiple substitutions must occur sequentially, because of the absence of recombination (Muller 1932). Each successive substitution therefore entails a further delay during which no other can occur.

Consider a favorable new mutation with selection coefficient S. For those mutations that are eventually fixed, the initial frequency of the new mutant is, on the average, $\sim 1/N_e$, and the increase from this initial frequency to some high frequency is essentially deterministic. In our experimental cultures N_e is $\sim 3.3 \times 10^7$, and, from the step model, S is ~ 0.1. During the time that its numerically dominant predecessor undergoes 250 doublings, the initially rare favored mutant ($W = 1.1$) would undergo 275 doublings. These 25 extra doublings are sufficient

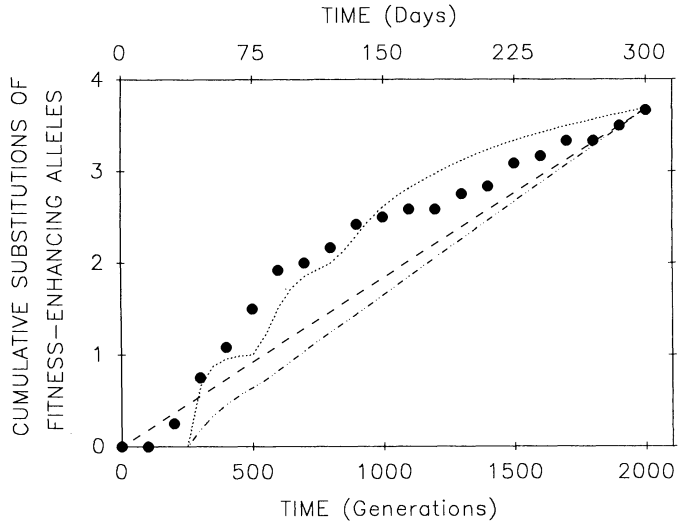

FIG. 6.—Theoretical and observed patterns of substitution of fitness-enhancing alleles. *Filled circles*, cumulative number of substitutions per population inferred from the data by using the step model (table 1). The three lines give expectations based on different theoretical models of the substitution process; models A (*dashed line*), B (*dots and dashes*), and C (*dotted line*) are described briefly in the legend to fig. 5 and in greater detail in the text.

to allow the favored mutant to increase by a factor of $2^{25} \cong 3.4 \times 10^7$ relative to its predecessor; the resulting ratio of the two clones would be about 1:1 and the corresponding frequency of the favored mutant would be ~0.5, so that the favored mutant would then appreciably influence the population's mean fitness (fig. 2). (Subsequent substitutions of comparable selective value should entail comparable delays, owing to the "interference" between successive clones.) By contrast, if $S = 0.01$, then we would expect a delay of roughly 2,500 generations before a new mutant appreciably influenced its population mean fitness, unless the mutation rate to the favored genotype, u, was much greater than $1/N_e$. But if $u \gg 1/N_e$, such that 30 or more substitutions with $S = 0.01$ could occur within 2,000 generations, then we might not expect any noticeable delay before mean fitness began to increase. Thus, the observed delay of 200–300 generations before mean fitness begins to increase noticeably (fig. 1) is consistent with selection coefficients for the favored alleles of ~0.1, which suggests that the number of such substitutions that have occurred, on the average, is about three or four, and not 30 or 40 as would follow from model A.

To generate the trajectories of mean fitness and its variance using model B, we have again assumed that the expected number of substitutions during 2,000 generations is 3.67 and that the associated selection coefficients are all 0.1. As shown above, $S = 0.1$ also implies a delay, D, equal to 250 generations for each substitution. One can solve model B by using an iterative procedure, as described in the Appendix. The value that is obtained for k is $\sim 4.1 \times 10^{-3}$ per generation, which corresponds to $u \cong 6.2 \times 10^{-10}$ per cell generation.

The resulting trajectory of the variance of mean fitness between populations is shown in figure 5. Note that the rate of divergence that is predicted by model B is much less than that predicted by model A, although it is still somewhat greater than the observed trajectory of variance of mean fitness between populations. This reduction in the expected variance can be understood as follows. Each substitution of a favorable allele entails a delay during which no other substitution of an allele of comparable selective value can occur. For each such substitution during a finite interval, the time available for any additional substitution is correspondingly reduced, which creates a negative feedback and thus maintains a lower variance of the number of substitutions between populations.

Therefore, model B can account for the initial delay that is observed in the trajectory of mean fitness, which model A cannot. Model B also generates less variance of mean fitness between populations than does model A, which is again consistent with our observations. However, model B, like model A, assumes no limit to the number of equally likely and equally advantageous mutations that can eventually occur. Hence, neither model A nor model B can account for the declining rates of substitution (fig. 6) and fitness increase (fig. 1) with time.

Model C, Poisson process with delays and a finite number of classes of advantageous mutations.—Models A and B both assume that the likelihood of a fitness-enhancing mutation during any particular time interval is independent of the number of beneficial alleles that have been fixed previously. In model C, we consider the effects of having a finite number of fitness-enhancing mutations, or classes of like mutations.

Two key properties of model C are as follows. First, as particular adaptive substitutions are "used up," the likelihood of further fitness-enhancing substitutions is reduced. Hence, the expected rate of increase in mean fitness diminishes with time. This clearly is consistent with the trajectory of mean fitness observed in the experimental populations (fig. 1; table 1). Second, it is implicitly assumed that all populations eventually converge to the same fitness state. Hence, variance between populations of mean fitness must ultimately be transient and cannot be sustained indefinitely. This second property appears superficially to be inconsistent with the significant variance of mean fitness between populations that remains even after 2,000 generations (table 2).

To determine whether this apparent inconsistency is real, we examined various sets of parameters for model C for those that might be consistent with both the trajectory for mean fitness and the pattern of variation in mean fitness between populations. Model C is much more complex than model B, in that the different classes of adaptive mutations may vary in their underlying rates and associated selection coefficients. However, the solution of model C is simplified somewhat if only the mutation rates for each class, but not the associated selection coefficients, are varied (see Appendix). In calculating the probability that two or more substitutions occur in a specified number of generations, one must further stipulate that each mutation belongs to a different class. Once again, we sought a solution to our model that gave rise to an expectation of 3.67 substitutions after 2,000 generations, with each selection coefficient equal to 0.1. Unlike models A or B, for which there is only a single solution that fulfills these criteria, there are

many combinations of mutation rates for four or more classes of favorable mutation that can fulfill these criteria for model C. Moreover, some combinations of mutation rates fit quite well *both* the trajectory for the between-population variance in mean fitness (fig. 5) and the temporal pattern of substitutions (fig. 6), given the substantial statistical uncertainties in the empirical estimates. (The particular combination of mutation rates that is shown in these figures has three classes of mutation with $u = 1 \times 10^{-9}$ per cell generation, and two other classes of mutation with $u = 7 \times 10^{-11}$ per cell generation.) Thus, the failure to discern the asymptotic decay in between-population variance during 2,000 generations, even as the trajectory of mean fitness becomes progressively more shallow, is not inconsistent with model C.

It should be emphasized that we are not presenting any particular set of parameters as the "best fit" to the data. Rather, our point is that a simple model, one that assumes eventual convergence of all mean fitness trajectories to the same final value, is qualitatively and quantitatively consistent with all of the major features of the data. Model C thus provides a credible *null* hypothesis with respect to the *alternative* hypothesis that the replicate populations are diverging from one another in a fitness landscape that contains multiple peaks of varying heights.

DISCUSSION

We can summarize the main findings of our study as follows. (1) Mean fitness increased in all 12 evolving populations, apparently owing to only a few substitutions that produced an average increase in fitness of ~37% relative to their common ancestor. There was no evidence for nontransitive fitness interactions among clones. (2) The rate of increase in mean fitness diminished over time, suggesting an approach to a plateau for mean fitness. (3) Statistically significant within-population genetic variance of fitness was not detected over the course of the experiment. However, the observed low levels of within-population variance conform to theoretical expectations for successive selective substitutions of the magnitude observed. (4) Statistically significant between-population variance of mean fitness was detected throughout the course of the experiment. The observed levels of between-population variance were consistent with a model having a finite number of nonepistatic classes of beneficial mutation. This model *predicts* that between-population variance in mean fitness will ultimately be transient, although the 2,000 generations of evolution reported here do not clearly indicate whether between-population variance will be transient or sustained.

Patterns of Adaptation and Divergence

In the introduction to this article, we set out three alternative scenarios for the adaptation and divergence of replicated populations evolving in parallel environments. According to the first scenario, all populations eventually converge to the same stable adaptive peak, although stochastic variation in the time of origin of particular classes of beneficial alleles among the replicate populations gives rise to transient divergence. Under the second scenario, ecological interactions among clones are such that the adaptive surface itself is unstable, owing to nontransitive

fitness effects. The third scenario predicts that replicate populations approach separate peaks of unequal mean fitness on a stable adaptive surface, such that variance of mean fitness between populations is sustained indefinitely.

In order to distinguish among these alternatives, we monitored three fundamental and interdependent variables: mean fitness, the genetic variance of fitness within populations, and the genetic variance of mean fitness between populations. On the basis of comparisons of the observed trajectories of these three variables with expectations derived from population-genetics theory, we have been unable to reject the first and simplest scenario—that is, eventual convergence of the replicate populations to a single, stable adaptive peak. We consider further some of the evidence and its limitations, below.

Trajectory of mean fitness.—An important feature of the trajectory for mean fitness over 2,000 generations is that it becomes relatively more shallow with time (fig. 1; table 1). This type of trajectory is expected if there are a finite number of classes of advantageous mutations, each class with its own mutation rate and selection coefficient. Assuming that each class of mutations can produce only one selective substitution, then there will be a tendency toward progressively longer waiting times between selective substitutions. If there is variation in mutation rate among these classes, then those classes that have higher rates will tend to generate selective substitutions earlier in the evolutionary process. Variation in selection coefficients among classes of mutations can also cause progressively longer intervals between selective substitutions, as the likelihood of eventual fixation is reduced for less advantageous mutants and the period required for their substitution is longer. Therefore, although mean fitness appeared to reach some sort of plateau by the end of 2,000 generations, we cannot rule out further substitutions of favorable mutations that occur at lower rates or have smaller selection coefficients than those observed thus far.

A model in which there are a finite number of classes of advantageous mutations thus suffices to explain the diminishing rate of increase in mean fitness, even if one assumes strictly additive effects of selective substitutions. A qualitatively different kind of explanation for this trend is epistatic interaction among loci, such that later substitutions have progressively less effect due to patterns of gene expression (Hartl et al. 1985).

Transitivity.—Paquin and Adams (1983) observed that the fitness of yeast populations decreased over the course of laboratory evolution, relative to the ancestral strain. Their result apparently reflects nontransitive fitness interactions among clones. In our study of bacterial populations, we observed no instances in which mean fitness was lower than that of the ancestor. However, it is still possible that nontransitive effects could have arisen during particular selective substitutions without giving rise to a net decrease in mean fitness relative to the ancestor. This too was tested for each of the 12 evolving populations by using the pair of consecutive isolates that had given the greatest apparent drop in mean fitness. On retesting, these drops were not statistically reproducible, which suggests that there is a relative absence of nontransitivity in the evolution of the populations studied here.

Why is there this difference between our results and those of Paquin and Adams

(1983)? Both studies were conducted in glucose-limited medium, although their populations were maintained in chemostats, whereas our populations were maintained by serial transfer. It can be shown mathematically by using simple chemostat models, that two clones cannot stably coexist on a single limiting resource, and that the "winning" clone is the one that holds the equilibrium resource concentration to the lowest level (Stewart and Levin 1973; Hsu et al. 1977; Hansen and Hubbell 1980; Tilman 1982; Waltman 1983; Lenski and Hattingh 1986). It follows, therefore, that relative fitnesses should be strictly transitive when clones are competing for a single limiting resource in a chemostat. Hence, there must have been other environmental factors, such as growth inhibitors (see, e.g., Chao and Levin 1981) or secondary resources produced as metabolic by-products (see, e.g., Helling et al. 1987), that caused the nontransitive fitness effects seen by Paquin and Adams.

As there was no evidence for nontransitive fitness interactions in our study, we see no need to invoke the effects of metabolic by-products in determining relative fitnesses of the clones isolated during the 2,000 generations of evolution. It is perhaps relevant that the serial transfer mode of culture propagation and the low glucose concentration used in our experiment may reduce the importance of metabolic by-products, as their concentrations are expected to be proportional to population density. We conclude that, in our study, the increased fitness of the evolved clones was most likely due to a shortened lag phase on transfer to fresh medium, a higher maximal growth rate, a greater affinity for the limiting glucose as it becomes depleted, a reduced rate of mortality during stationary phase, or some combination of these adaptations.

Fisher's fundamental theorem.—A basic expectation from evolutionary theory is that the rate of change in mean fitness should be proportional to the genetic variance of fitness itself (Fisher 1930; Nagylaki 1977). Therefore, in this study, we explicitly estimated the variance of fitness between clones within populations in order to examine its correspondence to the rate of change in mean fitness. We observed a large increase in mean fitness (~37% during 2,000 generations, including ~26% between generations 200 and 600). However, we were unable to measure statistically significant within-population genetic variance of fitness, despite 48 ANOVAs involving 12 populations at four points in time (144 clones and 288 fitness assays total). Nevertheless, the theoretical expectations for the within-population genetic variance of fitness, based on the trajectory of mean fitness, were statistically indistinguishable from the empirical estimates. For these bacterial populations, it is evidently much easier to discern genetic variance from the response to selection than by direct measurement of the variance.

The nature of the adaptive landscape.—Some of the most difficult questions in evolutionary theory concern the adaptive landscape, which is typically visualized as multiple peaks of varying heights separated along only one or two genetic axes (e.g., Wright 1982). But how smooth or rugged are real adaptive surfaces, and on what scale? Under what circumstances can evolving populations shift between alternative peaks?

At face value, our results suggest a smooth adaptive surface, with little ten-

dency for replicate populations to diverge toward alternative adaptive peaks. Although we did detect statistically significant variance of mean fitness between populations (table 2), its magnitude was no greater than that expected theoretically, given stochastic variation in the timing of a series of sequential, nonepistatic, selective substitutions, with a single fitness peak.

Proponents of the view that evolution cannot be as simple as it seems to be from our results have at least three lines of argument available against this conclusion. First, we have data for "only" 2,000 generations. While these data are entirely consistent with a simple model in which all populations are converging on a state of equal fitness (model C; figs. 5, 6), a corollary of this model—that variance of mean fitness between populations is asymptotically eliminated—remains unproved. A more definitive test of the "single peak" model may require many thousands more generations of observation. This model can be falsified if the level of between-population variance of mean fitness remains significant indefinitely, even in the absence of further increases in mean fitness. In such a case, it would become impossible to identify parameters for model C consistent with the trajectories of *both* mean fitness and its variance.

Second, one of the limitations in the data analyzed here is that they are confined to fitness. We may be missing between-population heterogeneity in the characters that underlie fitness (see, e.g., Hoffmann and Cohan 1987; Lenski 1988*a*). It is possible that the populations have evolved to similar mean fitnesses, but by different physiological adaptations to the culture environment. Some concrete possibilities for characters that may have differentiated over populations include the length of the lag phase upon transfer to fresh medium, the maximal growth rate, the affinity for limiting glucose as it becomes depleted, and the mortality rate during stationary phase. We are now assaying these characters for the 12 populations, as sampled at 2,000 generations. Even if these fitness-related characters are not differentiated in the standard culture environment, it may be possible to distinguish the populations from each other on the basis of differences under other environmental conditions. However, variation in characters unrelated to fitness in the environment in which the populations have evolved cannot be taken as evidence for *adaptive* divergence.

Third, this entire experiment has been confined to the simplest possible context for evolutionary divergence: an initially uniform set of populations evolving in a uniform environment. A more incisive test of the potential for divergence during adaptation to a uniform environment might use populations with different antecedent evolutionary histories (cf. Cohan and Hoffmann 1986). An example of such a test would be to take the 12 evolving populations from this experiment and place them in a novel environment, replicating each of these populations. In this way, it might be possible to start evolution at different points on the adaptive surface, so as to sample more extensively the range of possible trajectories. Going further, one might initiate evolution with populations that have different selective histories (see, e.g., Lenski 1988*b*), perhaps arising from culture in different media or at different temperatures (see, e.g., Bennett et al. 1990). Regardless of the scale of this variation in initial genetic composition, if the adaptive surface is

indeed smooth, with a single peak, then all the replicate populations should eventually converge on the same adaptations. The present experiment is only a first step on the adaptive landscape.

ACKNOWLEDGMENTS

R E.L. dedicates this paper to the memory of his friend, Ralph Evans, Jr. We thank K. Dao, J. Mittler, D. Seaborg, and M. Travisano for assistance with the experiments. We are also grateful to A. Bennett, F. Cohan, J. Felsenstein, S. Frank, D. Hudson, J. Mittler, and M. Travisano for especially helpful discussions of this work. The study was supported by a grant from the National Science Foundation (BSR-8858820) to R.E.L.

APPENDIX

Solutions for Models

ITERATIVE SOLUTION FOR MODEL B

Model B assumes that all substitutions have equal selection coefficients, as does model A. Model B also includes a time delay, D, that is associated with each adaptive substitution. During this interval, no other substitution can occur owing to linkage disequilibrium and the resulting "interference" between clones (Muller 1932). It is possible to solve model B for the expected mean and variance of the number of substitutions as a function of time, provided that one is given D and the expected number of substitutions, $\mathrm{E}(B(T))$, for one particular point in time, T.

From equation (8a), k is the substitution rate of favorable mutations per generation, according to a simple Poisson process, without any delays or interference between clones. Hence, k must also be the rate of *appearance* of favorable alleles that would eventually be fixed if there were no interference between clones. Similarly, kt is the expected number of mutations during t generations that would eventually be substituted if there were no interference between clones.

To solve model B iteratively, one makes an initial estimate of k. By using this estimate, the Poisson distribution can be used to obtain the relative frequencies of all possible outcomes necessary for computing the mean and variance of the number of substitutions after a specified number of generations. For example, let $D = 250$ generations and $T = 600$ generations, so that the number of substitutions possible is zero, one, or two. To obtain the probability that exactly two substitutions occur in 600 generations, one must realize that $2D = 500$ generations are taken up by the delays associated with the substitution process. One then computes the probability that two or more mutations occur in the remaining $T - 2D = 100$ generations, $p(\geq 2, 100)$; from the Poisson, $p(\geq 2, 100) = 1 - e^{-100k} - 100ke^{-100k}$. One next computes the probability that exactly one substitution occurs in 600 generations, but now also recognizing that for one substitution to be manifested by 600 generations, one or more mutations destined for fixation would have to occur in $T - 1D = 350$ generations. Again using the Poisson, one obtains $p(\geq 1, 350) = 1 - e^{-350k}$, which is the probability that one or more substitutions would occur in 600 generations. The probability that exactly one substitution occurs in 600 generations is given by $p(\geq 1, 350) - p(\geq 2, 100)$. For zero substitutions to occur in 600 generations would require that zero mutations occur in $T - 1D = 350$ generations; the probability of this is given by $p(0, 350) = e^{-350k}$.

For the initial estimate of k, one can thus obtain the probability of all outcomes necessary to compute the mean number of substitutions during T generations. Therefore, by iteratively increasing or decreasing k, one can identify a value that uniquely fulfills the expectation, $\mathrm{E}(B(T))$.

After the appropriate value for k has been determined, one can use the same computational algorithms to obtain the relative frequencies of all possible numbers of substitutions at any other time t. From these relative frequencies, one then obtains the mean and variance between populations of the number of substitutions as a function of time. One can also use equation (8a) to back-calculate from k the rate of favorable mutation, u, provided that the other parameters are also known.

ITERATIVE SOLUTION FOR MODEL C

Model C assumes a finite number of classes of beneficial mutation. Model C is much more complex than model B, in that these different classes of mutation may vary in their underlying rates and associated selection coefficients. If selection coefficients are allowed to vary, then the delay, D, associated with the substitution process must also vary among the different classes of mutation. Variation in selection coefficients would also create asymmetries with respect to the "interference" between clones. However, obtaining solutions for model C is simplified considerably if only the mutation rates for each class, but not their associated selection coefficients, are allowed to vary. In this case, solution of model C is similar to that for model B, except that it becomes necessary to keep track of each of the different multilocus genotypes.

For example, let $D = 250$ generations and $T = 600$ generations (as for model B above). Now let us also assume that there are only two distinct classes of beneficial mutation. We can define k_1 and k_2 as the rate of *appearance* of favorable alleles belonging to the two respective classes that would eventually be fixed if there were no interference between clones and no limit to the number of substitutions. Similarly, k_1t and k_2t are the expected number of mutations during t generations that would eventually be substituted if there were no interference between clones and no limit to the number of substitutions.

To obtain the probability that exactly two substitutions occur in 600 generations, one must compute the probability that one or more mutations of each type occur in $T - 2D = 100$ generations, $p(\geq 1, \geq 1, 100)$, which equals $(1 - e^{-100k_1})(1 - e^{-100k_2})$. One next computes the probability that at least one mutation of either type occurs in $T - 1D = 350$ generations; this is given by $p(\geq 1, 0, 350) + p(0, \geq 1, 350) + p(\geq 1, \geq 1, 350) = 1 - p(0, 0, 350) = 1 - (e^{-350k_1})(e^{-350k_2}) = 1 - e^{-350(k_1 + k_2)}$. The probability that exactly one substitution of either class occurs is then obtained by subtracting $p(\geq 1, \geq 1, 100)$ from this quantity. For zero substitutions to occur by 600 generations would require that zero mutations occur in $T - 1D = 350$ generations; the probability of this outcome is given by $p(0, 0, 350) = e^{-350(k_1 + k_2)}$.

One proceeds iteratively by varying the k_i in order to satisfy the constraints imposed by $E(B(T))$ and D. However, for model C, there may be many different combinations of mutation rates for a finite number of classes of beneficial mutation that will satisfy these contraints. Among these many different parameter sets, it may be possible to identify some that give a good fit to the observed trajectories of mean fitness and the variance of mean fitness between populations.

LITERATURE CITED

Atwood, K. C., L. K. Schneider, and F. J. Ryan. 1951. Periodic selection in *Escherichia coli*. Proceedings of the National Academy of Sciences of the USA 37:146–155.

Bennett, A. F., K. M. Dao, and R. E. Lenski. 1990. Rapid evolution in response to high-temperature selection. Nature (London) 346:79–81.

Bouma, J. E., and R. E. Lenski. 1988. Evolution of a bacteria/plasmid association. Nature (London) 335:351–352.

Carlton, B. C., and B. J. Brown. 1981. Gene mutation. Pages 222–242 *in* P. Gerhardt, ed. Manual of methods for general bacteriology. American Society for Microbiology, Washington, D.C.

Chao, L., and B. R. Levin. 1981. Structured habitats and the evolution of anticompetitor toxins in bacteria. Proceedings of the National Academy of Sciences of the USA 78:6324–6328.

Chao, L., B. R. Levin, and F. M. Stewart. 1977. A complex community in a simple habitat: an experimental study with bacteria and phage. Ecology 58:369–378.

Christiansen, F. B. 1975. Hard and soft selection in a subdivided population. American Naturalist 108:157–166.

Clark, B. C., P. R. Shelton, and G. S. Mani. 1988. Frequency-dependent selection, metrical characters and molecular evolution. Philosophical Transactions of the Royal Society of London B, Biological Sciences 319:631–640.

Cohan, F. M., and A. A. Hoffmann. 1986. Genetic divergence under uniform selection. II. Different responses to selection for knockdown resistance to ethanol among *Drosophila melanogaster* populations and their replicate lines. Genetics 114:145–163.

Cohan, F. M., A. A. Hoffmann, and T. W. Gayley. 1989. A test of the role of epistasis in divergence under uniform selection. Evolution 43:766–774.

Crow, J. F. 1968. The cost of evolution and genetic loads. Pages 165–178 *in* K. R. Dronamraju, ed. Haldane and modern biology. Johns Hopkins University Press, Baltimore.

Dykhuizen, D. E., and D. L. Hartl. 1983. Selection in chemostats. Microbiological Reviews 47:150–168.

Feller, W. 1957. An introduction to probability theory and its applications. Vol. I. Wiley, New York.

Felsenstein, J. 1971. On the biological significance of the cost of gene substitution. American Naturalist 105:1–11.

Fisher, R. A. 1930. The genetical theory of natural selection. Clarendon, Oxford.

Gaines, S. D., and W. R. Rice. 1990. Analysis of biological data when there are ordered expectations. American Naturalist 135:310–317.

Ginzburg, L. R. 1981. Bimodality of evolutionary rates. Paleobiology 7:426–429.

Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection. V. Selection and mutation. Proceedings of the Cambridge Philosophical Society 23:838–844.

Hansen, S. R., and S. P. Hubbell. 1980. Single-nutrient microbial competition: qualitative agreement between experimental and theoretically forecast outcomes. Science (Washington, D.C.) 207:1491–1493.

Hartl, D. L., D. Dykhuizen, and A. M. Dean. 1985. Limits of adaptation: the evolution of selective neutrality. Genetics 111:655–674.

Helling, R. B., C. N. Vargas, and J. Adams. 1987. Evolution of *Escherichia coli* during growth in a constant environment. Genetics 116:349–358.

Hoffmann, A. A., and F. M. Cohan. 1987. Genetic divergence under uniform selection. III. Selection for knockdown resistance to ethanol in *Drosophila pseudoobscura* populations and their replicate lines. Heredity 58:425–433.

Hsu, S. B., S. P. Hubbell, and P. A. Waltman. 1977. A mathematical theory for single-nutrient competition in continuous cultures of microorganisms. Society for Industrial and Applied Mathematics Journal of Applied Mathematics 32:366–383.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Kleinbaum, D. G., and L. L. Kupper. 1978. Applied regression analysis and other multivariate methods. Duxbury, North Scituate, Mass.

Lenski, R. E. 1988*a*. Experimental studies of pleiotropy and epistasis in *Escherichia coli*. I. Variation in competitive fitness among mutants resistant to virus T4. Evolution 42:425–433.

———. 1988*b*. Experimental studies of pleiotropy and epistasis in *Escherichia coli*. II. Compensation for maladaptive effects associated with resistance to virus T4. Evolution 42:433–440.

Lenski, R. E., and S. E. Hattingh. 1986. Coexistence of two competitors on one resource and one inhibitor: a chemostat model based on bacteria and antibiotics. Journal of Theoretical Biology 122:83–93.

Lenski, R. E., and B. R. Levin. 1985. Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. American Naturalist 125:585–602.

Levin, B. R., F. M. Stewart, and L. Chao. 1977. Resource-limited growth, competition, and predation: a model and experimental studies with bacteria and bacteriophage. American Naturalist 111:3–24.

Luria, S. E., and M. Delbrück. 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511.

Moran, P. A. P. 1962. The statistical processes of evolutionary theory. Oxford University Press, Oxford.

Moser, H. 1958. The dynamics of bacterial populations maintained in the chemostat. Carnegie Institution, Washington, D.C.

Mosteller, F., and R. E. K. Rourke. 1973. Sturdy statistics: nonparametrics and order statistics. Addison-Wesley, Reading, Mass.

Muller, H. J. 1932. Some genetic aspects of sex. American Naturalist 66:118–138.

Nagylaki, T. 1977. Selection in one- and two-locus systems. Springer, New York.

Paquin, C., and J. Adams. 1983. Relative fitness can decrease in evolving populations of *S. cerevisiae*. Nature (London) 306:368–371.

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution 43:223–225.

Robertson, T., F. T. Wright, and R. L. Dykstra. 1988. Order restricted statistical inference. Wiley, New York.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry. 2d ed. W. H. Freeman, San Francisco.

Stewart, F. M., and B. R. Levin. 1973. Partitioning of resources and the outcome of interspecific competition: a model and some general considerations. American Naturalist 107:171–198.

Tilman, D. 1982. Resource competition and community structure. Princeton University Press, Princeton, N.J.

Wallace, B. 1968. Topics in population genetics. Norton, New York.

Waltman, P. 1983. Competition models in population biology. Society for Industrial and Applied Mathematics, Philadelphia.

Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proceedings of the Sixth International Congress of Genetics 1:356–366.

———. 1938. Size of population and breeding structure in relation to evolution. Science (Washington, D.C.) 87:430–431.

———. 1982. Character change, speciation, and the higher taxa. Evolution 36:427–443.

———. 1988. Surfaces of selective value revisited. American Naturalist 131:115–123.