**Proceedings of the 2007 IEEE Symposium on
Artificial Life (CI-ALife 2007)**

Predicting Fitness Effects of Beneficial Mutations in Digital Organisms

Haitao Zhang[1]and Michael Travisano[1]

[1]Department of Biology and Biochemistry, University of Houston, Houston TX 77204-5001

Abstract-**Evolutionary adaptation can be viewed as two separate processes. The first process is the origin of new beneficial mutations. The second process is the fixation of some of those beneficial mutations by natural selection. Instead of statistical descriptions of adaptive changes, evolutionary theory is now focusing on predicting fitness effects of beneficial mutations in response to selection. While population genetics has provided an extensive body of theory to predict evolutionary changes, it is often difficult to predict evolution since many factors interact to affect the selective coefficients necessary for prediction. Here, we provide experimental data to study the ability of predicting evolutionary changes by using digital organisms (ALife program). We are concerned with how the dynamics of adaptation and diversification are determined by sequential fixation of beneficial mutations. More specifically, we are interested in the rates of fitness changes in populations and the distribution of fitness effects of beneficial mutations. Our results confirm the diminishing return of the rates of fitness increase. A step model provides a best fit to fitness trajectory of populations. The diminution in the rates of fitness increase is due to both a decrease in step sizes and an increase in waiting times. The distribution of fitness effects among beneficial mutations is nearly exponential except for some small fitness changes of beneficial mutations.**

I. INTRODUCTION

Fisher [1] and Muller [2] viewed the evolution of asexual populations as the sequential fixation of beneficial mutations. Fisher's geometric model [1] brought forward a simple question: which phenotypic changes are more likely to be beneficial, small ones or large ones? By showing that mutations with larger effects have a rapidly declining possibility of being favorable, he concluded that small mutations have larger propensities to be beneficial so that small phenotypic changes are therefore the primary substrate for adaptation. Fisher's adaptive walks also implied that early substitutions tend to have larger phenotypic effects than later ones. Nevertheless, when the genetic drift of mutations with small effects is taken into account, the conclusion could be different. Kimura [3] first pointed out that mutations with intermediate effects may play an important role in adaptation. Later, Gillespie [4-5] proposed a mutational landscape model by studying DNA sequence evolution. He argued that the distribution of fitness effects among beneficial mutations is always exponential and always has the same mean, which is independent of present wild-type alleles. Interestingly, Gillespie's model gave a result similar to Fisher's geometric model: larger-effect substitutions appear earlier than small-effect ones. This is often referred to as the pattern of diminishing returns [1, 6-7].

Although a diverse range of organisms including plants and animals are employed in evolution experiments, most studies on the dynamics of adaptive evolution have been done using microorganisms, due to their short generation times and large population sizes. Previous studies have shown that populations of microorganisms evolve quickly at first and that the rates of adaptation decline over time [8-10]. Divergence of populations and species from one another is often viewed as a consequence of adaptation. However, even for replicate populations that are founded from a common ancestral clone and are selected in identical environments, the dynamics of divergence may be very complicated [11]. Replicate populations, starting with no variation for fitness, are expected to diverge due to the stochastic processes associated with the appearance and fixation of beneficial mutants, but may or may not subsequently converge [8]. In spite of numerous studies of adaptation and divergence using microbial populations, the general dynamics of variance in mean fitness between replicate populations remains an open question.

Several mathematical models have been proposed to describe the dynamics of adaptation during long-term evolution experiments. For example, a hyperbolic model was applied to describe adaptation in a 10,000-generation experiment with bacterial populations [9]. This model provided a better fit to the trajectory of cell size over time, compared to a linear regression model. A step model was also applied to the same populations in order to provide finer resolution to the first 3000 generations of the cell size trajectory [12]. Using the step model, the authors argued that punctuated evolution results were due to sequential selection and fixation of beneficial mutations within the replicate populations. The application of step methods provided a much better understanding of adaptive dynamics than previous methods. However, these studies only showed descriptive data and only carried out very preliminary analysis using the step model. More replicates of these kinds of experiments and more precise measurements of parallel fitness changes are necessary to fully understand the punctuated dynamics. Using digital organisms, we perform long-term replicated evolution experiments to answer the following question: can we predict the rate of fitness change due to fixation of beneficial mutations and the distribution of fitness effects of such mutations? Our approach is to track the long-term fitness trajectories during adaptation and divergence of digital organisms in a simple environment. This study offers the most precise monitoring yet of the dynamics

1-4244-0701-X/07/$20.00 ©2007 IEEE

of long-term evolution and distribution of fitness effects among beneficial mutations.

II. EXPERIMENTAL DESIGN AND METHODS

All experiments were conducted using Avida version 2.0b7 and run on a Mac G5 running OS 10.3 operating systems. We used the default settings on Avida unless otherwise indicated.

Model System: Digital organisms (e.g., Avida program) are computer programs that can self-replicate, mutate, compete and evolve by natural selection. They can be viewed as a domesticated form of computer viruses that compete for the use of the central processing unit (CPU), which is necessary for replication. Unlike other computer programming approaches using evolutionary optimization such as genetic algorithms, digital organisms explicitly replicate their genomes to reproduce and no optimal sequence is designated as the target. Avida is a well-developed software platform for research in computational evolutionary biology [13]. Avidian organisms have similar genetic architectures to asexual organisms (e.g., bacteria). Each Avidian organism has a circular genome made up of a sequence of individual instructions that are analogous to proteins. There are 26 possible instructions, and these can be altered from one to another by mutation. Several types of mutations can happen in Avidian organisms. Copy mutation is the most common, which is caused by an error in copying instructions from parents to offspring during reproduction.

There are many advantages to study evolutionary biology using digital organisms. It is faster and easier to do long-term selection using digital organisms, and we can observe the dynamic evolutionary process at a level that is unobtainable in traditional experiments. For example, we can track the fitness of digital organisms every generation and the experiments can be replicated hundreds of times, providing greater experimental precision and higher statistical accuracy. Recent studies using digital organisms have investigated long-term microevolution and macroevolution [14-18].

Adaptive Evolution: Twenty populations of digital organisms, founded by a single clonal ancestor, were propagated in parallel at the mutation rate of 0.75 per genome per generation in a simple environment for 100,000 updates (~10,000 generations). The genome length of digital organisms was kept at 100 instructions during selection. Each population evolved under the restriction of environmental carrying capacity (3,600 individuals). The selection pressure in the present study can be viewed as a simple novel environment, which only favors organisms with fast replication rate. Fitness was estimated every 100 updates (~10 generations) over the entire 100,000 updates (~10,000 generations) for analysis of fitness trajectories. Intervals of 1000 updates (~100 generations) were used to analyze the rates of fitness increase. Intervals of 500 updates (~50 generations) were used to fit the step model.

Index of Divergence versus Parallelism: The roles of adaptation, chance and history in promoting adaptation and divergence can be disentangled by using fitness data [19]. The effects of adaptation are measured as the fitness increase relative to the ancestral clone. The combined roles of chance and history are estimated as the square root of variance in fitness among replicate populations. Historical effects in this project are zero since all replicates are derived from a single ancestral clone. We use the square root of fitness variance (standard deviation) in order to compare the effects of adaptation with effects of chance and history in same unit [20]. Finally, an index of the relative extent of divergence versus parallelism is given by (1). Parallelism refers to the similarity of fitness between replicated populations, while divergence measures the average difference in fitness among populations.

$$I_x = \frac{\sqrt{Var_G}}{\left|\Delta w\right|} \tag{1}$$

where Var_G is the variance in fitness among genotypes and represents the effects of chance and history, and Δw is the average fitness change relative to the ancestral clone, indicating the effects of adaptation.

In our experiment, parallelism is defined strictly in terms of fitness. Since populations could achieve almost identical finesses via different genetic and phenotypic trajectories, parallel evolution in this experiment may not indicate that phenotypically similar types evolve. This index measures the average divergence in fitness among the evolved populations relative to the average fitness change from the ancestral state, and indicates whether the replicate populations evolved in a more parallel or a more divergent manner. A value of I_x less than one would suggest that the differences among the populations are smaller than the average fitness change from the ancestral state. Alternatively, a value of I_x greater than one would suggest that the evolution of the replicate populations is more divergent.

Rate of Adaptive Evolution: The rate of adaptive evolution at time point i is determined by the slope of fitness from point i to the last point. The equation (2) for the calculation of the rate of adaptive evolution is described below:

$$\frac{m_{all}X_{last} - m_{1:i}X_i}{X_{last} - X_i} \tag{2}$$

where m_{all} is the slope of all points, $m_{1:i}$ is the slope of the first i points, X_{last} is the x-axis value of the last point, and X_i is the x-axis value of the ith point.

Hyperbolic Model: Each fitness trajectory of the 20 replicate populations is fitted to a hyperbolic model by using the following function (3):

$$y = x_0 + \frac{at}{b+t} \tag{3}$$

where y is the fitness at time t, x_0 is the ancestral fitness, a and b are parameters in a specific case.

The Step Model (Punctuated Evolution): Each fitness trajectory of the 20 replicate populations is fitted to a step model (4) by using reduced isotonic regression procedures [21]. Isotonic regression, a nonparametric method, is a powerful statistical model when a dependent variable is monotonically related to an independent variable. The isotonic regression model is used to estimate a step-wise regression, which divides the description of n points into l groups ($l < n$); thus, this method yields l homogenous groups. The estimate for each group equals the average fitness (dependent variable) for points in the group; however, the number of groups under an isotonic regression model is often large. The problem can be solved by performing a backward elimination algorithm to combine adjacent groups that do not differ significantly under a partial *F*-test [21-22]. This procedure is called the reduced isotonic method.

$$y = c_0 + \sum_{i=1}^{n} c_i \cdot (if\ t \geq T_i) \qquad (4)$$

where y is the fitness at time t, c_0 is the ancestral fitness, c_i is the fitness increase during the ith step, T_i is the time at which the ith step occurs, and n is the number of steps used in a specific case.

III. RESULTS

Temporal Trajectory for Fitness of Populations

Strong responses to selection in the simple environment are observed during the course of evolution. Fig. 1 plots the fitness of each replicate population against time over 100,000 updates. The average fitness increased by about 78% relative to the ancestral clone. Individual fitness followed a similar trend in each replicate population: fitness increased over time with some fluctuation, though there was a slight decline in fitness at the beginning of selection for almost all populations. Visual inspection of Fig. 1 clearly indicates that fitness among all populations increased quickly at first and the rates of fitness increase subsequently slowed down. The fitness values reached their peak after adaptation for 50,000 updates and subsequently remained stable for the duration of the experiment. The twenty populations can be divided into three groups based on their peak values of fitness. The first group includes seventeen populations, which increased fitness dramatically and reached their peak values (~1.78) in the first 22,000 updates. A single population forms the second group, which evolved to its fitness peak at about 5,000-updates, and had a relatively low peak value (~1.56). The third group consists of two populations, whose fitness peak values are between 1.9 and 2.0. The emergence of three distinct groups of evolved populations suggests that the groups evolved by different paths and approached different fitness peaks in the adaptive landscape.

Temporal Trajectory for Index of Divergence versus Parallelism

The trajectory for the index of divergence versus parallelism (Fig. 2) shows a trend that is very similar to the trend in the variance. The index goes up quickly and reaches its peak value (~0.19) at 4,000-updates. It then goes down dramatically and approaches a stable non-zero value (~0.05) at about 40,000-updates. Since the value of the index is much smaller than 1, we conclude that the derived populations evolved in a highly parallel manner during the course of evolution.

Rates of Adaptive Evolution

We analyzed the trajectories for average rate of fitness among replicated populations increase (2) and its standard errors at 1,000 update intervals for the 20 replicate populations (Fig. 3). The rates of mean fitness increase observed here give supporting evidence to the inference from the genetic theory of adaptation and previous microbial evolution studies: populations in novel environments evolve quickly at first and the rate of adaptation slows down later. The initial rate of fitness increase is about 2.2 x 10^{-5} per update. The rate then drops dramatically in the first 20,000 updates and fluctuates with a long tail until it approaches zero. Interestingly, the values of standard error among 20 replicate populations are restricted to a narrow range and are nearly constant over the whole evolutionary process, which implies that each of 20 derived populations had a similar trend in adaptive speed despite the divergence among them.

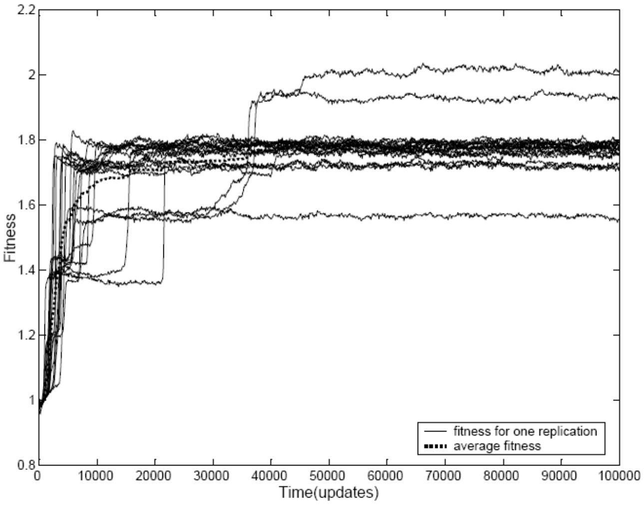

Fig. 1. Trajectories for fitness during 100,000 updates of evolution with 20 replicates in the simple environment. Fitness is estimated relative to an ancestral clone at 100 update intervals.

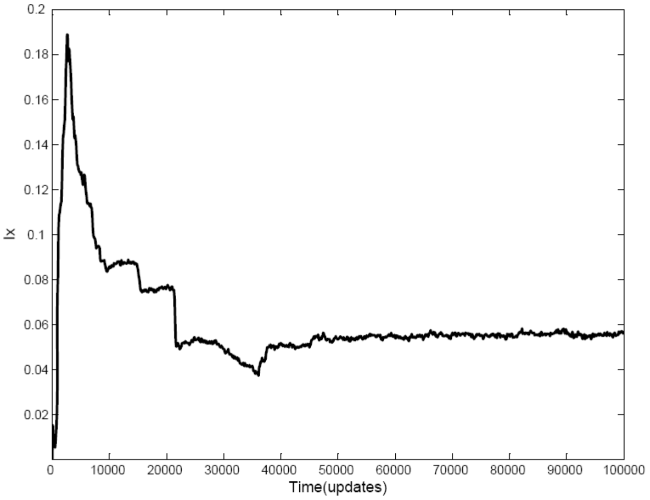

Fig. 2. Trajectory for index of divergence versus parallelism during 100,000 updates of evolution in the simple environment.

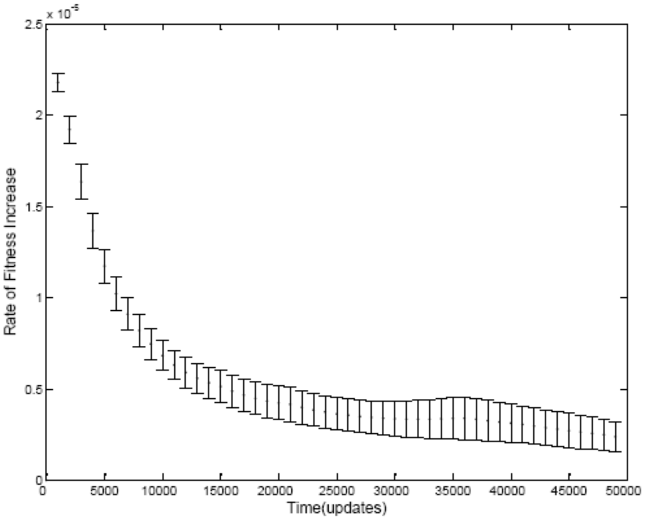

Fig. 3. Trajectory for rates of fitness increase among the 20 derived populations during 100,000 updates of evolution in the simple environment. Dots represent average rates of fitness increase estimated every 1,000 updates. Error bars show corresponding standard errors based on 20 populations.

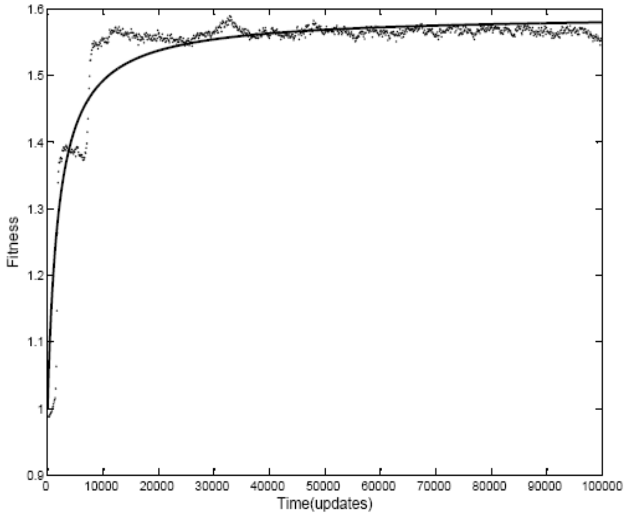

Fig. 4. Trajectory for fitness in one population during 100,000 updates of evolution in the simple environment. Dots are relative fitness of this population at 100 update intervals. The solid curve is the best fit of a hyperbolic model.

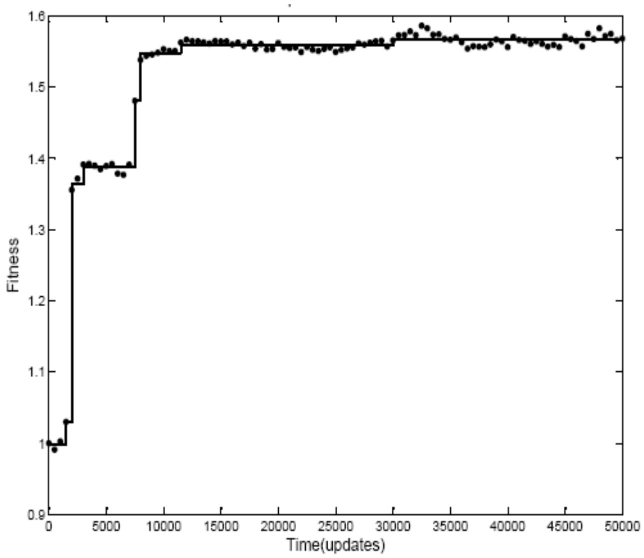

Fig. 5. Trajectory for fitness in one population during the first 50,000 updates of experimental evolution using a step model with eight steps. Dots are values of fitness estimated every 500 updates. Solid line shows the best fit of a step model.

The Hyperbolic Model

We first plot the fitness trajectory line for one case population during 100,000 updates of evolution in the simple environment (Fig. 4). The following analysis of prediction models is also based on the same population. Fitness increased quickly and approached a value of 1.4 at 4,000-updates after the clone was introduced into the selective environment. A temporary period of stasis then appears and is sustained for about 3,000 updates. A new fitness jump is observed at 7,000-updates and the peak fitness is observed near 8,000-updates. This fitness is sustained over the rest of the 100,000 updates. A hyperbolic model (3) fits the fitness trajectory of this population well (Fig. 4, $r = 0.9250$, $n = 1001$, $p < 0.001$). That is to say, the rate of fitness increase declines over time following an initial rapid rate, and it approaches an asymptotic value. Relative to a linear regression model with the initial fitness constrained to 1, a hyperbolic model provides a significantly better fit (partial $F = 39492$, $df = 2$ and 998, $p < 0.0001$). The data described by the hyperbolic model suggest rapid adaptation after the population was introduced in the simple environment, followed by an evolutionary stasis during the remaining part of the long-term evolution. We applied the hyperbolic model to the fitness data for each of the remaining 19 derived populations and found a good fit in every case. For all 20 populations, the improvement in fit achieved by using the hyperbolic model is highly significant ($p < 0.0001$), relative to the linear model.

Step Model (Punctuated Evolution)

For each derived population, we fitted its fitness trajectory for the first 50,000 updates using a step model (4). Since all populations reached their maximum fitness in the first half of the evolutionary experiment, we ignore the fitness data from the last 50,000 updates of evolution. A model with eight discrete steps provides a best fit to the fitness data of the population shown in Fig. 5. We determined the step model trend line by using the fitness data in 500 update intervals. Theoretically, we can capture one beneficial mutation in each step by using a small interval data (500 updates), although it is possible there is more than one mutation in some steps. Even so, the eight-step model for the population in Fig. 5 gives a significantly better fit than a hyperbolic model in the first 50,000 updates of evolution (partial $F = 676.19$, $df = 5$ and 93, $p < 0.0001$).

Each step is caused by the fixation of one or more beneficial mutations via natural selection. The magnitude of each step, generally referred to as step size, is the fitness increase between two adjacent steps, which reflects the fitness effects of the corresponding beneficial mutations. There are two types of waiting times for fixation of beneficial mutations. One is the number of generations required for a given beneficial substitution to fix within the population. The other is the generally longer time period required before the appearance of a beneficial mutation due to low overall rate of mutation and low probability that a given mutation is beneficial. The step model we use in this project only considers the second type of waiting time. Beneficial mutations are therefore assumed to spread and dominate the population very quickly, once they arise. For each of the 20 derived populations, the best-fit step model utilizes from four to fifteen discrete steps. The step model provides an excellent fit for all 20 populations. The improved fit provided by the step model for all of the derived populations, relative to a hyperbolic model, is highly significant ($p < 0.0001$).

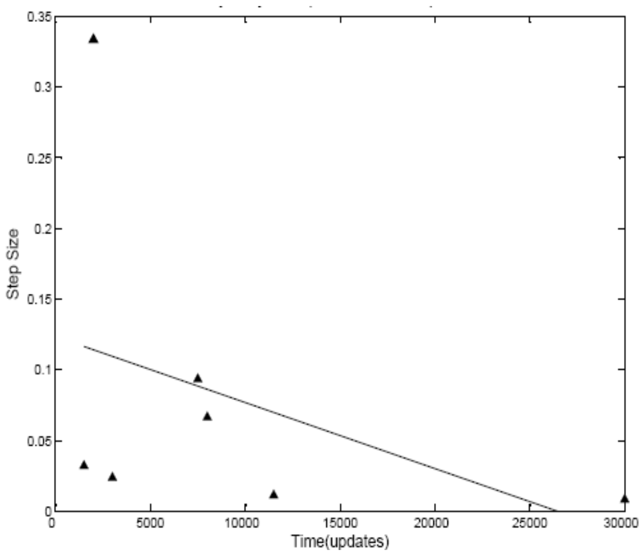

Fig. 6. Trajectory for step sizes against time scale in one population during 100,000 updates of evolution in the simple environment. Step sizes (filled triangles) are the fitness gain for corresponding steps (shown in Fig. 5). The trend line is the best fit of the linear regression model.

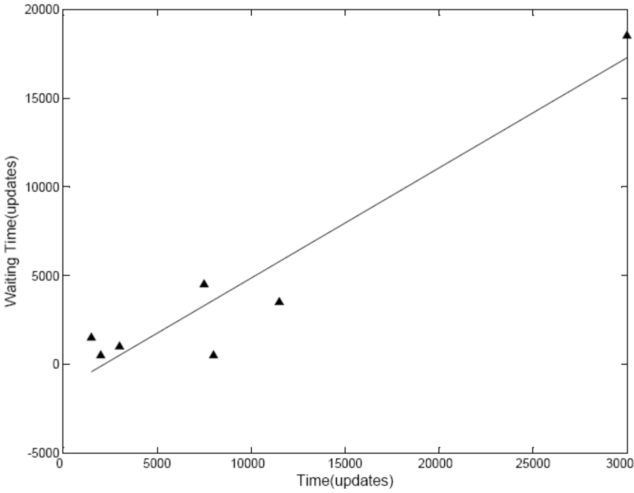

Fig. 7. Trajectory for waiting times against time scale in one population during 100,000 updates of evolution in the simple environment. Waiting times (filled triangles) are the time periods (updates) required for the appearance of beneficial mutations (shown in Fig. 5). The trend line is the best fit of the linear regression model.

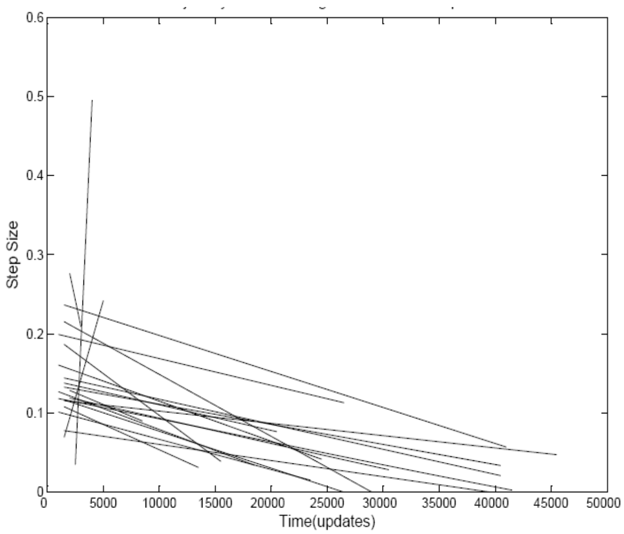

Fig. 8: Linear regression trend lines for step sizes over time in each of 20 populations during 50,000 updates of evolution in the simple environment. Step sizes are the fitness gain for corresponding steps due to fixation of beneficial mutations.

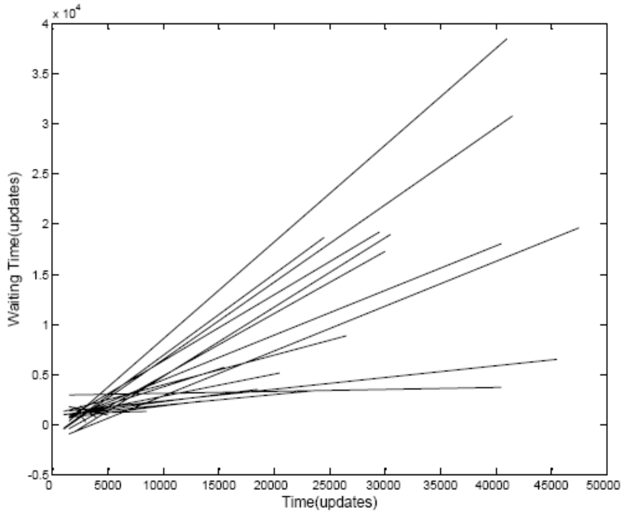

Fig. 9. Linear regression trend lines for waiting times over time in each of 20 populations during 50,000 updates of evolution in the simple environment. Waiting times are time periods required for the appearance of beneficial mutations.

Dynamics of Adaptive Evolution Based on Step Model

As already mentioned, the step model is a powerful method to analyze the rate of adaptive change in populations. Visualization of fitness lines in the step model suggests declining returns of the fitness effects of beneficial substitutions (Fig. 5). We may wonder whether the decline in the rates of fitness increase is primarily caused by a reduction in the step sizes or an increase in the waiting times between successive steps. The observed diminishing return of the rate of fitness increase may be complicated since it reflects an interaction between step sizes and waiting times. To determine the factors for the diminishing rate of fitness increase, we plotted the trajectory for step sizes and waiting times against time, and the corresponding trend lines for these points in one population (Fig. 6 and Fig. 7). The trend line in Fig. 6 suggests a reduction in average step sizes during the evolution. However, a positive relationship between the waiting time and the time scale is found in Fig. 7. Accordingly, the observed decline in the rates of fitness increase is evidently caused by both the reduction in step sizes and the increase in waiting times for this population.

To determine if the previous conclusion based on a single population holds for other replicate populations, we plotted all linear trend lines for the twenty replicate populations together (Fig. 8, Fig. 9). The lengths of trend lines are different, due to the various numbers of steps for each population under the step model. Fig. 8 plots the regressed lines for step sizes in the 20 replicate populations over time for 50,000 updates of evolution. Among the 20 trend lines for step size, 18 of them have negative slopes, which indicate a reduction in the step sizes. For the trajectories of waiting time, most regressed lines have an upward slope with only three exceptions (Fig. 9). These findings clearly show that the diminution in the rate of fitness increase is the result of both a decrease in step sizes and an increase in waiting times.

Distribution of Fitness Effects among Beneficial Mutations

Gillespie's mutational landscape model concluded that the distribution of fitness effects among beneficial mutations is nearly exponential in spite of wild-type alleles. To test this theory, we collected all data of fitness effects among beneficial mutations which are estimated from 149 steps sizes by using the step model in the twenty replicate populations. We then plotted the distributed curve for fitness effects of beneficial mutations by using a histogram chart in Excel (Fig. 10). This histogram chart shows that the fitness effects of beneficial mutations fixed during adaptation are nearly exponentially distributed when we ignore some small values of fitness increase. The distributed curve is highly skewed (asymmetrical). Actually, the skewed histogram is typical in many important biological properties, such as numbers of species in a genus, life spans of species, etc [23].

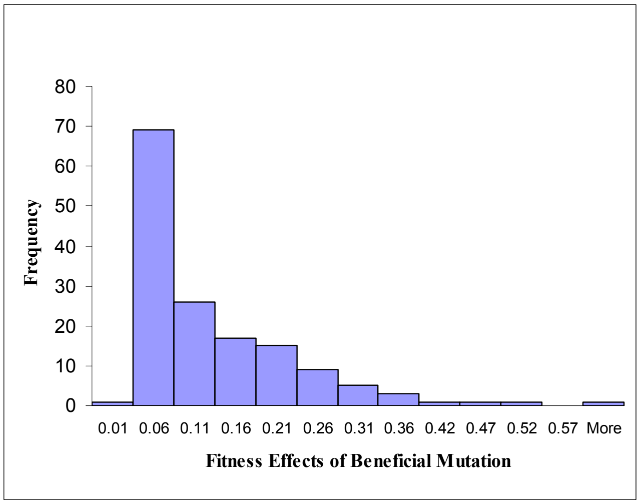

Fig. 10. Distribution of fitness effects among beneficial mutations in 20 derived populations during 100,000 updates of evolution in the simple environment. 149 values of fitness increase for beneficial mutations are estimated from step sizes by using a step model.

IV. DISCUSSION

Genetic Causes for Diminishing Rates of Adaptive Evolution

The results indicate a decline in rate of fitness increase over time based on both a hyperbolic model and a step model. The genetic and adaptive mechanisms underlying the diminishing return of fitness increase may be very complicated. Several theories have been developed to explain the declining rates. For example, one hypothesis proposed that there exist a finite number of groups of advantageous substitutions, where each group has its own mutation rate and selection coefficient [8]. The author further argued that those groups with higher mutation rates and higher selection coefficients are likely to generate mutations with larger effects earlier during adaptive evolution. The likelihood of further beneficial mutations is reduced as more and more adaptive mutations are used up; therefore, the expected rates of fitness gain are reduced with time. This hypothesis could be used to explain our results. Although the genomic mutation rate is kept constant in our experiments, the mutation rates of beneficial substitutions could still be variable in different regions of the genome.

Epistatic interactions among genes have also been proposed to explain the diminishing rate of adaptive evolution [24]. That is to say, the likelihood of beneficial alleles arising during any particular time interval is dependent on the beneficial mutations that have been fixed previously. Due to antagonistic epistatic interaction between later beneficial substitutions and earlier ones, the rate of adaptive evolution slows down. This line of reasoning raises a basic question: does a subset of currently beneficial mutations constrain subsequent evolution? Three groups with different fitness peaks having different long-term fitness trajectories shown in Fig. 1 clearly imply constraints due to earlier substitutions. For example, only two populations reach the highest peak fitness, while most populations (n = 18) apparently have no chance of reaching the highest peak value. This might be a correlated effect of adapting quickly in the first part of evolution. Strong epistatic interactions between earlier substitutions and later ones may constrain the effects of later substitutions. The population with the lowest fitness in our experiment may have used up its beneficial mutations as it did not improve for the last 80% of the selection, although running the experiment for even longer could possibly have resulted in a fitness improvement.

Parallelism versus Divergence

Values of the index of parallelism versus divergence far less than one in our experiment indicate that the derived populations evolved in parallel, relative to the divergence observed during the evolution. This is reasonable since the independently replicated populations were selected in an identical environment. Another possibility is that the simple environment setting in this experiment with no externally complex selection pressure makes the evolution highly parallel. Populations which adapted in different genetic pathway could still approach the similar fitness in the simple environment. Although the divergence among populations is small, relative to the overall mean adaptive response, it is not transient and tends to remain even after 100,000 updates of evolution. We believe that the divergence among populations is caused either by epistatic interactions between earlier substitutions and later ones or by epistatic constraints of earlier substitutions on the availability of later ones. These possibilities underscore the importance of the role of chance in adaptive evolution, because the order of substitutions can affect the evolutionary process significantly.

NK models [25] introduced the idea that different sequence spaces might generate different local optima. Moreover, the smoothest landscape featured a single global optimum, while a rugged landscape featured many local optima so that adaptive walks tended to approach to the nearest local optimum. The final sequence space may not evolve from a transiently dominant genotype. For instance, we could imagine the fixation process of beneficial mutations: a single genotype with one beneficial mutation sweeps through the population and becomes the dominant genotype; however, this dominant genotype may subsequently crash due to the chance effects (mutation and genetic drift). Another possibility is the population may be dominated by an intermediate genotype that is not favored by natural selection; therefore, the next beneficial mutation to sweep through the population may come from the second fit genotype. That is to say, derived populations from an identical ancestral clone could adapt to different peaks on the adaptive landscape by chance.

Punctuated Evolution

Our data clearly demonstrate a pattern of punctuated evolution for fitness during the first part of evolution. Long periods of ‘stasis’ are followed by relatively brief spurts of rapid fitness change, which is consistent with punctuated equilibrium theory. Natural selection of rare beneficial substitutions may be responsible for the punctuated increase in fitness [12]. In our experiments, derived populations are selected in the simple environment from an ancestral clone that can only reproduce asexually. Mutations are thus the only source of genetic change during adaptive evolution. Due to the low mutation supply rate, a population would spend many generations (stasis) waiting for a rare beneficial substitution. On the other hand, the fixation of the beneficial mutations is very fast relative to the long waiting time for the appearance of beneficial mutations. Trajectory for fitness lines in the step model suggests declining returns of the fitness effects of beneficial substitutions. In their remarkable work, Elena and his colleagues [12] concluded that the observed decrease in the rate of fitness increase is due to an increase in the waiting time between successive steps. Our results, however, demonstrate that both a decrease in the step sizes and an increase in the waiting time are responsible for the diminishing return of rates of the fitness increase. Furthermore, step sizes play a major role in the declining rate of fitness change in the first half of the experimental evolution, while the contribution of waiting time is made larger and larger in the second half, partly because many adaptive substitutions had been used up.

We found fitness data obtained at different time intervals would affect the shape of the step model. The trajectory line for step model shown in Fig. 5 is plotted by using the fitness data at 500 update intervals. We have chosen fitness data at three different time intervals (100 updates, 500 updates and 1,000 updates) to match the step model. The shapes of trajectory lines are different (not shown). Furthermore, it will affect the estimation of adaptive rate of evolution. Our data show that it is hard to disentangle the steps if we choose the fitness data at 100 update intervals. The fitness data at 500 update intervals were finally chosen for analysis of punctuated evolution.

Distribution of Fitness Effects among Beneficial Mutations

The primary obstacle to develop a genetic theory of adaptation is the unknown distribution of fitness effects among beneficial mutations. Consequently, there has been a long-standing interest to study this topic; however, this approach has proven to be very difficult because of two main problems: how to capture the rare beneficial substitutions in an experimental system, and how to measure the fitness effects of small beneficial mutations.

Recently, studies of experimental microbial evolution have focused on uncovering the distribution of beneficial effects, but their findings are distorted by both stochasitc loss and clonal interference [26-27]. On the other hand, theoretical work has also been focused on this topic. For example, Gillespie’s mutational landscape model studies the adaptation of DNA sequences by using extreme value theory [4-5]. These studies show that the distribution of beneficial effects should always be exponential, independent of other biological complexities. To test these theories, we collected sample data for the fitness effects of beneficial mutations estimated from step sizes by using a step model. In this way, we could detect and measure the fitness effects of rare beneficial mutations, including those with small effects. As we mentioned earlier, each step could represent the fixation of one or more beneficial mutations. Considering the small time interval (500 updates) in our experiment, the step sizes can roughly be used to estimate the fitness effects of beneficial mutations. Our findings indicate that the fitness effects of beneficial mutations are nearly exponentially distributed except for some small values of step sizes. Interestingly, Orr's simulation results [6] also found some small values of beneficial effects, but he ignored these small sizes as noise in the subsequence analysis. Because our data are obtained from a step model, we believe these small sizes represent the effects of real beneficial mutations.

ACKNOWLEDGMENTS

We thank members of Dr. Travisano's laboratory for many helpful comments on the manuscript. HT Zhang especially thanks Michael Quance for his great assistance and comments.

REFERENCES

[1] R. A. Fisher, The genetical theory of natural selection, Oxford University Press, Oxford, 1930.
[2] H. J. Muller, "Some genetic aspects of sex," The American Naturalist, vol. 66, pp.118-138, 1932.
[3] M. Kimura, The neutral theory of molecular evolution, Cambridge University Press, Cambridge, UK, 1983.
[4] J. Gillespie, "Molecular evolution over the mutational landscape," Evolution, vol. 38, pp.1116-1129, 1984.
[5] J. Gillespie, The Causes of Molecular Evolution, Oxford University Press, Oxford, 1991.
[6] H. A. Orr, "The population genetics of adaptation: the distribution of factors fixed during adaptive evolution," Evolution, vol. 52, pp.935-949, 1998.
[7] N. H. Barton, and P. D. Keightley, "Understanding quantitative genetic variation," Nature Reviews Genetics, vol. 3, pp.11-21, 2002.
[8] R. E. Lenski, M. R. Rose, S. C. Simpson, and S. C. Tadler, "Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations," The American Naturalist, vol. 138, pp.1315-1341, 1991.
[9] R. E. Lenski, and M. Travisano, "Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations," Proc. Natl. Acad. Sci. U.S.A., vol. 91, pp. 6808-6814, 1994.
[10] K. K. Holder and J. J. Bull, "Profiles of adaptation in two similar viruses," Genetics, vol. 159, pp.1393-1404, 2001.
[11] F. M. Cohan, and A. A. Hoffmann, "Genetic divergence under uniform selection. II. Different responses to selection for knockdown resistance to ethanol among Drosophila melanogaster populations and their replicate lines," Genetics, vol. 114, pp.145, 1986.
[12] S. F. Elena, V. S. Cooper, and R. E. Lenski, "Punctuated evolution caused by selection of rare beneficial mutations," Science, vol. 272, pp.1802-1804, 1996.
[13] C. Ofria, and C. O. Wilke, Avida user's manual (http://dllab.caltech.edu/avida), 2003.
[14] G. Yedid, and G. Bell, "Microevolution in an electronic microcosm," The American Naturalist, vol. 157, pp. 465-487, 2001.
[15] G. Yedid, and G. Bell, "Macroevolution simulated with autonomously replicating computer programs," Nature, vol. 420, pp.810-812, 2002.
[16] R. E. Lenski, C. Ofria, T. C. Collier, and C. Adami, "Genome complexity, robustness, and genetic interactions in digital organisms," Nature, vol. 400, pp.661-664, 1999.
[17] C. O. Wilke, J. L. Wang, C. Ofria, R. E. Lenski, and C. Adami, "Evolution of digital organisms at high mutation rate leads to survival of the flattest," Nature, vol. 412, pp.331-333, 2001.
[18] R. E. Lenski, C. Ofria, R. T. Pennock, and C. Adami, "The evolutionary origin of complex features," Nature, vol. 423, pp.139-144, 2003.
[19] M. Travisano, J. A. Mongold, A. F. Bennett, and R. E. Lenski, "Experimental tests of the roles of adaptation, chance, and history in evolution," Science, vol. 267, pp.87-90, 1995.
[20] F. Vasi, M. Travisano, and R. E. Lenski, "Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment," The American Naturalist, vol. 144, pp.432-456, 1994.
[21] M. J. Schell, and B. Singh. "The reduced monotonic regression method," JASA, vol. 92, pp.128-35, 1997.
[22] D. G. Kleinbaum, and L. L. Kupper, Applied regression analysis and other multivariate methods, Duxbury, North Scituate, Mass, 1978.
[23] D. M. Raup, Extinction: bad gene or bad luck? W.W. Norton & Company, Inc, 1991.
[24] D. L. Hartl, D. E. Dykhuizen, and A. M. Dean, "Limits of adaptation: the evolution of selective neutrality," Genetics, vol. 111, pp.655-674, 1985.
[25] S. Kauffman, "Towards a general theory of adaptive walks on rugged landscapes," J. Theor. Biol., vol. 128, pp.11-45. 1987.
[26] M. Imhof, and C. Schlotterer, "Fitness effects of advantageous mutations in evolving Escherichia coli populations," Proc. Natl. Acad. Sci. U.S.A., vol. 98, pp.1113-1117, 2001.
[27] D. E. Rozen, J. A. G. M. De Visser, and P. J. Gerrish, "Fitness effects of adaptations in microbial populations," Curr. Boil., vol. 12, pp.1040-1045, 2002.