This is to certify that the
dissertation entitled

ADAPTATION AND SPECIALIZATION IN BIOLOGICAL
AND DIGITAL ORGANISMS

presented by

ELIZABETH ANNE OSTROWSKI

has been accepted towards fulfillment
of the requirements for the

Doctoral degree in Zoology

Richard E. Lenski

Major Professor's Signature

17 October 2005

Date

ADAPTATION AND SPECIALIZATION IN BIOLOGICAL AND DIGITAL ORGANISMS

By

Elizabeth Anne Ostrowski

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Zoology

2005

ABSTRACT

ADAPTATION AND SPECIALIZATION IN BIOLOGICAL AND DIGITAL ORGANISMS

By

Elizabeth Anne Ostrowski

The transition from generalist to specialist may involve two components. On one hand, it may involve adaptations that improve fitness in a subset of the environment. However, specialization may also entail the evolutionary loss other functions or abilities, resulting in narrow niche breadth. I used an experimental evolution approach to distinguish between two competing hypotheses for the evolution of specialization. One hypothesis is that specialization results from antagonistic pleiotropy, which arises when mutations that improve fitness in one environment cause fitness declines in other environments. The other hypothesis is that specialization results from mutation accumulation, whereby relaxed selection on unused functions allows mutations to accumulate in the genes encoding them, leading to their eventual loss. Using two distinct model systems, the bacterium *Escherichia coli*, and the digital evolution platform, *Avida*, I evaluated the contributions of these two mechanisms to the evolution of specialization.

In Chapter 1, replicate populations of *E. coli* evolved in a glucose-limited environment long enough to acquire their first beneficial mutation. I looked for evidence of antagonistic pleiotropy by examining the effects that these mutations had on fitness in other, unselected environments. Pleiotropy was common, such that most mutations had detectable fitness effects in other resources, but it was also generally positive. In addition, positive pleiotropic effects were often correlated, such that larger improvements

in one environment were associated with larger improvements in another environment. In Chapter 2, I sequenced candidate loci to determine the genetic bases of these mutations, uncovering a total of 21 mutations in 5 loci. In most cases, different substitutions in the same locus were phenotypically more similar to one another than they were to substitutions in other loci. In other cases, different substitutions in the same locus were associated with unique phenotypic effects.

In Chapter 3, I describe the evolution of specialization in digital organisms—self-replicating computer programs that compete, adapt and evolve. I founded replicate populations with generalist organisms, which could perform a variety of logic computations, and examined their evolution in environments where only a single computation, EQU, provided organisms with increased energy. Evolved populations often lost unselected functions, but the extent of these losses depended greatly on the ancestor. Some functions increased in performance, and these functions often exhibited high genetic overlap with the EQU function, suggesting that genetic architecture was an important component of niche breadth reductions. In Chapter 4, I investigated the consequences of genetic overlap for the ability of pairs of functions to evolve independently. I evolved populations in environments where one function was rewarded and the other was punished, or where one function was rewarded and the other evolved only as a correlated response. Despite strong positive correlated responses in some environments, most functions were capable of evolving independently. In one environment, I found evidence of multiple adaptive peaks. I examined two different hypotheses about evolution on rugged adaptive landscapes to explain the failure of populations evolving in this environment to reach the higher adaptive peak.

ACKNOWLEDGMENTS

This work could not have been completed without the support and guidance of my co-advisors, Drs. Richard Lenski and Charles Ofria. Rich provided generous financial support and resources, at levels uncommon for graduate students to receive. I am indebted to him for giving me the opportunity to work in his laboratory, as well as for the training and direction I received while there. As a mentor, Rich struck a careful balance between encouraging independence, while ensuring that I continued to make progress. I appreciate most of all that he held me to such high standards of work, while allowing me the freedom to accomplish it in my own way, in my own time.

I am indebted to Dr. Charles Ofria for taking me on as a graduate student at a point when I could not as much as open a terminal at a UNIX station. Charles taught me to think more critically and quantitatively. I may have baffled him at times with my inability to think mathematically, but he was always patient with his explanations and willing to start from the beginning. I see tremendous improvement in my computational skills as a direct result of having been a member of Charles' lab, and I leave here with entirely new areas of biology open to me.

I thank my committee members, Jeff Conner, Kay Gross, and Andy Jarosz for their support and advice on my thesis, as well as their willingness to take on an advisory role on this strange research with digital organisms. Janis Antonovics, Michael Hood, Peter Kareiva, Deborah Roach, Doug Taylor, and Henry Wilbur provided tremendous support and guidance during my undergraduate years and beyond. Peter Kareiva taught me that good experiments are not necessarily hard experiments, that clear thinking and

communication are contributions in and of themselves, and above all else, that science should be fun. Janis Antonovics encouraged me to be inquisitive and to think deeply about concepts. I will never forget Doug Taylor showing me a chestnut tree, then a chestnut blight canker, and then telling me that *maybe* (just maybe!) he would come back in six weeks to see how I was doing. Doug taught me that sometimes you are simply going to sink or swim.

I am indebted to a fantastic group of friends and colleagues who have supported me over the years. Cindy Wei often provided an ear, as well as many warm meals on days when I could not face my empty refrigerator. I thank Eva-Maria Muecke for her unflinching support and kindness. Bob Woods, Dule Misevic, and Kristina Hillesland were not only wonderful labmates and friends, but also the closest thing I had to family in Michigan. Danny Rozen has been a supportive friend and colleague over the years. I thank Chris Marx and Susi Remold for their insight into how to navigate the Ph.D. and beyond, as well as the many members of the Lenski lab over the years for providing such a supportive and stimulating laboratory environment.

I thank my parents, Peter and Carol Ostrowski, for their support over the years, for coming out to Michigan as often as they did, and for providing me passage home all those times when I really needed out. Finally, I owe an enormous debt of gratitude to Tim Cooper for being my labmate, my next-door neighbor, my colleague, and my partner—as well as for single-handedly bearing the brunt of all bad days. I continue to marvel at how I ever got along without him.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

PLEIOTROPIC EFFECTS OF BENEFICIAL MUTATIONS IN *ESCHERICHIA COLI*

The form and extent of pleiotropy is central to many theories in evolutionary biology, including the evolution of specialization (Cooper and Lenski 2000; Futuyma and Moreno 1988; Jaenike 1990), senescence (Rose 1991), and limits to adaptation (Barton and Keightley 2002; Otto 2004). The importance of pleiotropy was emphasized by Fisher, who formalized its role in his geometric model of adaptation (Fisher 1958). Fisher's model has two main assumptions: first, that pleiotropy is common, such that most mutations affect many or all of the traits under selection; and second, that the rate of environmental change is slow, such that organisms evolve towards a single, slow-moving optimum. As a consequence of the high "dimensionality" associated with adaptation, Fisher reasoned that only mutations of small effect would be adaptive.

Fisher's basic idea that pleiotropy is a common property of mutations has spurred other models that address the kinds of variation that are most likely to contribute to adaptation. For example, Lande (1983) formulated a model in which adaptation can be driven by numerous mutations of small effect or a few mutations of large effect. An important assumption of this model is that large mutations generally have large deleterious pleiotropic effects, whereas small mutations have smaller or no such effects. Unfortunately, evidence to support this assumption is lacking. As Orr and Coyne (1992) emphasize, not only does it require that highly beneficial mutations have deleterious pleiotropic effects, but the magnitude of these effects must generally outweigh the magnitude of the selected benefit. As they state:

> "Lande's micromutational theory requires that large mutations be *disproportionately* worse than smaller mutations. We simply have no information here. We do not know, for example, whether mutations adding four bristles to a fly are more than four times as harmful as mutations adding only a single bristle. Moreover, characteristics of deleterious mutations may tell us little about favorable ones. Even if large mutations are less fit than small ones, it does not follow that large mutations with a favorable primary effect suffer more deleterious effects than mutations with smaller primary effects."

In this study, we aim to test these suppositions raised by Orr and Coyne by examining the form and extent of pleiotropy associated with beneficial mutations in *Escherichia coli*. These mutations have a favorable primary effect: they increase fitness in a glucose-limited environment. Using these mutants, we examine the consequences of these mutations for fitness in novel resources. Specifically, we address the following pair of questions: (i) Do mutations that improve fitness in one resource environment have pleiotropic effects on fitness in alternative resource environments? In particular, will

improvements in one environment usually entail becoming worse in others? (ii) Do pleiotropic effects correlate in magnitude with the advantage in the primary selected environment? For example, will larger improvements entail larger trade-offs?

Several aspects of *E. coli* biology make it a good system for studying adaptation. Bacteria are easily cultured in the laboratory and can be stored in a non-growing state (a -80°C freezer), enabling direct comparison of evolved and ancestral strains. Short generation times and large population sizes can be achieved in the laboratory, such that adaptation occurs on observable time scales. Finally, replicate populations can be started from a single isogenic ancestor, so that multiple "random samples" of evolution from a single starting point can be obtained. This aspect allows us to collect a library of different, spontaneous beneficial mutations and to quantify their fitness effects relative to the ancestor.

Previous work has suggested that there is extensive pleiotropy associated with resource use in this system. For example, Remold and Lenski (2001) examined the fitness effects of random insertion mutations in novel resource and temperature environments and found that many of these mutations exhibited differential fitness effects across resource environments. Travisano *et al.* (1995) found that some populations evolved for 2,000 generations in glucose had improved fitness in maltose, but others had reduced fitness. The proposed explanation for this variability was that the populations had fixed beneficial mutations that differed in their pleiotropic effects on fitness in maltose. On other resources, populations showed either consistently increased fitness (NAG and mannitol),

or consistently decreased fitness (galactose and melibiose), suggesting that pleiotropy may be common, but that it differs in sign depending upon the resource in question (Travisano and Lenski 1996). Finally, Cooper and Lenski (2000) examined the consequences of 20,000 generations of adaptation to a glucose-limited environment for the ability to catabolize alternative resources. Replicate populations often showed parallel reductions of function, suggesting that the pleiotropic effects of beneficial mutations led to narrower catabolic niche breadth.

These studies have supported a general role for pleiotropy with regard to resource use, but two elements of these studies make it difficult to infer how common pleiotropy is for individual beneficial mutations. In the first study, collection of the mutations was random with respect to fitness (most were, in fact, deleterious) and produced by transposon insertions, meaning that they were disproportionately likely to be knock-out mutations; it is not clear how representative this kind of variation is with respect to adaptation. And whereas the long-term evolution experiments suggest that antagonistic pleiotropy underlies losses of function, it is not clear whether these findings were the result of many weakly pleiotropic mutations, or whether they were instead driven by one or a few mutations with highly pleiotropic effects. In general, there is little empirical data on the distribution of pleiotropic effects associated with beneficial mutations, despite the importance of this information to evolutionary theory (Otto 2004; Wingreen et al. 2003). Thus, in the current study, we seek a more direct assessment of the distribution of pleiotropic effects by examining individual beneficial mutations.

Materials and Methods

We used a collection of 30 genotypes that were independently evolved from a single ancestral genotype in an environment where glucose was the limiting carbon source (Rozen et al. 2002). These mutants were collected in such a way as to ensure that, in all likelihood, they each contain exactly one beneficial mutation. For each of these 30 mutants, we determined its fitness relative to the ancestor in glucose (the "selective" environment) and in 5 other resources (the "novel" environments).

Choice and Significance of Novel Resources

The novel resources we use vary in the extent to which their uptake and catabolism differs from that of glucose (Travisano and Lenski 1996). Two of the novel resources, NAG and mannitol, use the phosphoenolpyruvate transferase system (PTS) for transport across the inner membrane of the *E. coli* cell, which is also used for glucose transport. Resources that are transported using this system are phosphorlyated upon entry into the cell (Postma et al. 1996; Travisano and Lenski 1996). The non-resource-specific enzymes HPr and EI pass the phosphoryl group to the various resource-specific (EII) transporter proteins. By contrast, the three other novel resources we use—maltose, galactose, and melibiose—are non-PTS resources and use resource-specific permeases for transport into the cell. PTS enzymes are known to inhibit directly the uptake and metabolism of non-PTS resources, resulting in the preferential use of glucose even in the presence of other resources. In studies of populations evolved for 2,000 generations in a

glucose-limited medium, these resources showed a mixture of correlated responses ranging from primarily positive to primarily negative (Travisano and Lenski 1996).

Bacterial Strains and Culture Conditions

All genotypes used in this study were derived from the same *E. coli* B ancestor (Lenski et al. 1991). This ancestor contains no plasmids and is strictly asexual. In this system, different genotypes can be distinguished by the Ara marker, which denotes the ability (Ara^+) or inability (Ara^-) to catabolize L(+)arabinose. Genotypes with the Ara^- or Ara^+ marker state form red or white colonies, respectively, when plated on tetrazolium arabinose indicator agar. This color difference allows us to track the relative proportion of two genotypes (with opposite marker states) when we plate mixed cultures on indicator agar during competition assays.

In the current study, all mutants evolved from an ancestral strain that carried the Ara^+ marker. When assessing relative fitness, the evolved Ara^+ genotypes were competed in mixed cultures against the Ara^- variant of the ancestral strain. Previous studies have shown that the Ara^+ and Ara^- variants of the ancestor are neutral in the same glucose-limited culture conditions that were used in the current study (Lenski et al. 1991; Travisano et al. 1995). Moreover, we performed our own control competitions between the Ara^- and Ara^+ ancestral variants in glucose, as well as in the five novel resources. The results confirm that this marker difference is neutral in all resources used in the experiment (data not shown). While this requirement is not strictly necessary, it means

that competitions between Ara^+ evolved genotypes and the Ara^- variant of their ancestor can be assumed to be equivalent to competitions between the evolved genotypes and their Ara^+ (true) ancestor, with the marker serving only to allow the competitors to be distinguished.

Collection of Mutants

Collection of the mutants was described in detail in Rozen et al. (2002). Briefly, thirty replicate populations were founded with both Ara^+ and Ara^- variants of the ancestor in the following ratios (Ara^+:Ara^-): 1:100, 1:10, 1:1, 10:1, and 100:1. The populations were propagated daily by transferring 0.1 mL of the previous day's culture to 9.9 mL of fresh glucose medium. This 100-fold dilution permits a 100-fold daily increase in population size, equivalent to ~6.6 (or $\log_2 100$) generations per day. While the populations were being propagated, they were also plated periodically in order to assess the relative proportions of the two marker types. A sustained deviation in the ratio of the two marker types was indicative of the rise of a beneficial mutation in the increasing population—at which point, transfers were ceased for that population and clones of both marker types were isolated and saved. No population evolved beyond 400 generations, regardless of whether such a deviation had been observed.

Samples to determine relative frequencies of both markers were taken every day for the first seven days and then every other day thereafter. Monitoring of the trajectories at such high resolution allowed rapid and reliable detection of deflections from initial ratios.

In cases where deflections occurred prior to 400 generations, clones from earlier time points—near the point of deflection—were used. The maximum duration of the experiment was limited to 400 generations to minimize the possibility of fixation of double mutations. This time scale was chosen based upon earlier work (Elena et al. 1996; Gerrish and Lenski 1998; Lenski et al. 1991) that provided estimates for beneficial mutation rates and expected times to fixation for populations evolving under identical conditions.

For the purposes of the current study, we independently verified that a beneficial mutation was indeed present in each of these clones (see "Statistical Analyses"). In addition, it is highly probable that each of these mutants contains only a single beneficial mutation, owing to the discrete shifts in relative abundance used to detect the mutations and the short duration of the experiment. A convenient aspect of our bacterial system is that reproduction is asexual and thus, multiple beneficial mutations must arise sequentially on a single genetic background. Isolation of the mutants immediately after a shift in relative abundance in the mixed population makes it unlikely that a second beneficial mutation would have had sufficient time to arise on the same genetic background. To verify this expectation, Rozen et al. (2002) used a maximum likelihood approach to estimate the rate of beneficial mutations, using empirical estimates of the selection coefficient of the mutations and the deviation in marker ratios at the time of collection. Using these estimates, they found that in 100 simulations of their experiment, none of the "winning" clones (those whose frequency was increasing at the time of collection) had obtained more than one mutation.

Fitness Assays

Each mutant's relative fitness was determined via head-to-head competition assays with its ancestor in glucose (the selective environment) and five novel resources. The competition assays were performed as described by Lenski et al. (1991). Briefly, both competitors were inoculated from the freezer into 9.9 mL of Luria Broth (LB) and allowed to grow overnight. Next, as a preconditioning step, each mutant was transferred (via a 10,000-fold dilution) into each of the media types to be tested—generally glucose and one of the novel resources—and allowed to grow and acclimate for an additional 24 hours at 37°C. In all cases, Davis minimal medium was used, supplemented with 25 μg/mL of whichever carbon source was being tested.

The competition began when an equal volume (0.05 mL) of each competitor's pre-conditioned culture was transferred into a single test tube containing 9.9 mL of fresh medium and the appropriate carbon source. The competitors were then allowed to grow together for 24 hours, during which time they competed for the same pool of limiting resource. The competition cultures were plated at the beginning and end of the 24-hour period in order to assess the initial and final densities of the two competitors. From these densities, we calculated each competitor's realized Malthusian parameter, that is, its net rate of population growth during the competition. Relative fitness was then calculated simply as the ratio of the realized Malthusian parameters of the two competitors (Lenski et al. 1991).

We ran the competition experiments as sets of complete blocks. Within each block, we competed all 30 mutants against the ancestor in glucose and in one other resource. We then replicated this complete-block design three times for each of the five novel resources. Using this design, we obtained three independent fitness estimates for each mutant in the five novel resources, for a total of 450 competitions (30 genotypes × 5 resources × 3 blocks). In addition, fitness in glucose was measured in every block, such that we had a total of 450 estimates of fitness in glucose (30 genotypes × 15 blocks).

Of these 900 estimates of relative fitness, experimental errors led to the loss of two values: one in glucose and one in galactose. One additional missing value was introduced when we discarded a fitness estimate in glucose that was an extreme statistical outlier. This estimate was over eight standard deviations above the mean estimate in glucose. The loss of these estimates resulted in a very slight imbalance in our data sets; F-tests must therefore be considered approximate.

Statistical Analyses

Inclusion of mutants in the study

To establish that all of the presumptive mutants did, in fact, carry beneficial mutations, we calculated a mean and standard error for the fitness of each mutant in the selective (glucose) environment. As a preliminary step, we used a conservative two-tailed t-test to ask whether a mutant's relative fitness differed significantly from a null-hypothesized value of 1.0. If we failed to detect a statistically significant difference, we then

performed a series of additional 6-day competitions against the reciprocally marked ancestor in glucose. These longer competitions permit resolution of very small differences in fitness, enabling us to distinguish between a mutation of small effect and the absence of a mutation. Those few mutants that could not be confirmed to have increased fitness even following these more sensitive tests were dropped from subsequent analyses.

Heterogeneity of fitness effects in the selective environment

To determine the heterogeneity of fitness effects in the selective environment, we performed an analysis of variance (ANOVA) based on fitness estimates of the mutants in glucose only. This analysis was performed in SAS using PROC GLM (SAS Institute 1999), with block and genotype coded as random effects. Because each block contained a single replicate of each genotype, a genotype × block interaction was not included in the model. In order to quantify the heterogeneity in fitness effects in glucose due to differences among genotypes, we calculated a variance component for the effect of genotype. This component was calculated as the difference between the mean squares for genotype and error, divided by the sample size.

Pleiotropic effects of mutations in novel resources

To determine whether a particular mutant had pleiotropic effects in novel resources, we used two-tailed t-tests to assess whether its relative fitness in each resource differed from

a null-hypothesized value of 1.0, which would indicate fitness equal to the ancestor. We then assessed the heterogeneity of pleiotropic effects as a function of genotype and resource by performing a two-way ANOVA. This ANOVA was calculated using PROC GLM in SAS, with genotype and resource treated as random and fixed effects, respectively. F-tests were constructed according to Scheffé (1959), where the random effect of genotype was tested over the mean square error, and resource was tested over the interaction of genotype × resource. We also ran five separate one-way ANOVAs to assess the effect of genotype separately for each novel resource (Appendix Table A1). Similarly, we ran separate one-way ANOVAs to examine the effect of resource for each genotype (Appendix Table A2). These additional analyses aided the interpretation of the main effects of genotype and resource.

Scaling of pleiotropic effects

To address whether pleiotropic effects were proportional to the direct effect, we calculated Pearson correlation coefficients between fitness in glucose and in each of the five novel resources. Prior to the analysis, we assessed normality of the data using a Shapiro-Wilkes test in SAS. The correlation coefficients and their p-values were then obtained using SYSTAT (2002).

Our pairwise experimental design, in which genotypes competed against the ancestor simultaneously in glucose and in one other resource, permitted two possible routes for calculating correlation coefficients. When comparing fitness effects of each mutant in glucose and another resource, we could limit our glucose estimate to the mean of the

three estimates that were measured concurrently with the three estimates in the novel resource. This method would use a mean fitness in glucose based on fewer replicates, but controlled for possible variability across blocks. Alternatively, we could use the fitness in glucose calculated as the average of all 15 estimates. This method improves the accuracy of the estimate, but potentially conflates differences across blocks. After performing the analysis both ways, we found that, in all cases, the two methods produced the same significant results. For clarity, we present the analysis where a genotype's fitness in glucose is based on all 15 estimates, which makes it easier to compare the fitness effects of mutants across resources. However, because we used the same glucose estimates in all five correlations, we applied a sequential Bonferroni correction (Rice 1989) to these correlation results.

Where there was a significant correlation between fitness in glucose and a novel resource, we also calculated the slope of a Reduced Major Axis (Model II) regression, according to the following formula (Sokal and Rohlf 1995):

$$b_{RMA} = b_{OLS} / r_{XY},$$

where b_{OLS} is the slope of the ordinary least squares regression, and r_{XY} is the correlation coefficient between the fitness values of mutants in glucose and the alternative resource. The correlation coefficients, slopes of the least squares regressions, and their standard errors were calculated in SYSTAT. Using these estimates, we calculated confidence intervals on b_{RMA} as described in Sokal and Rohlf (1995).

Results

Size and Heterogeneity of Fitness Effects in the Selective Environment

Examination of the fitness values of the mutants in the selective glucose environment revealed that 27 of the 30 mutants tested were significantly more fit than the ancestor. The results of the additional 6-day competitions confirmed that the remaining three mutants did not differ from the ancestor in fitness (data not shown). These three mutants were dropped from further analyses, owing to the lack of evidence that they carried a beneficial mutation.

The remaining 27 mutants had a mean fitness in glucose of 1.096 (Fig. 1, far left). Results of the analysis of variance also revealed a significant effect of genotype (Table 1, $P < 0.0001$), indicating that these mutants are heterogeneous in the selective environment. There is also a significant block effect (Table 1, $P < 0.0001$), which indicates some fitness variation due to uncontrolled temporal differences in the environment (e.g., incubator temperature) across replicates. However, our pairwise experimental design, in which each genotype was simultaneously competed against the ancestor in glucose and one other resource in each block of the experiment, allows us to control for this variation across blocks. Finally, the variance component for genotype (calculated as the difference in the mean squares for genotype and error, divided by the sample size) is equal to 0.0007, and the square root of this estimate is 0.0264. In other

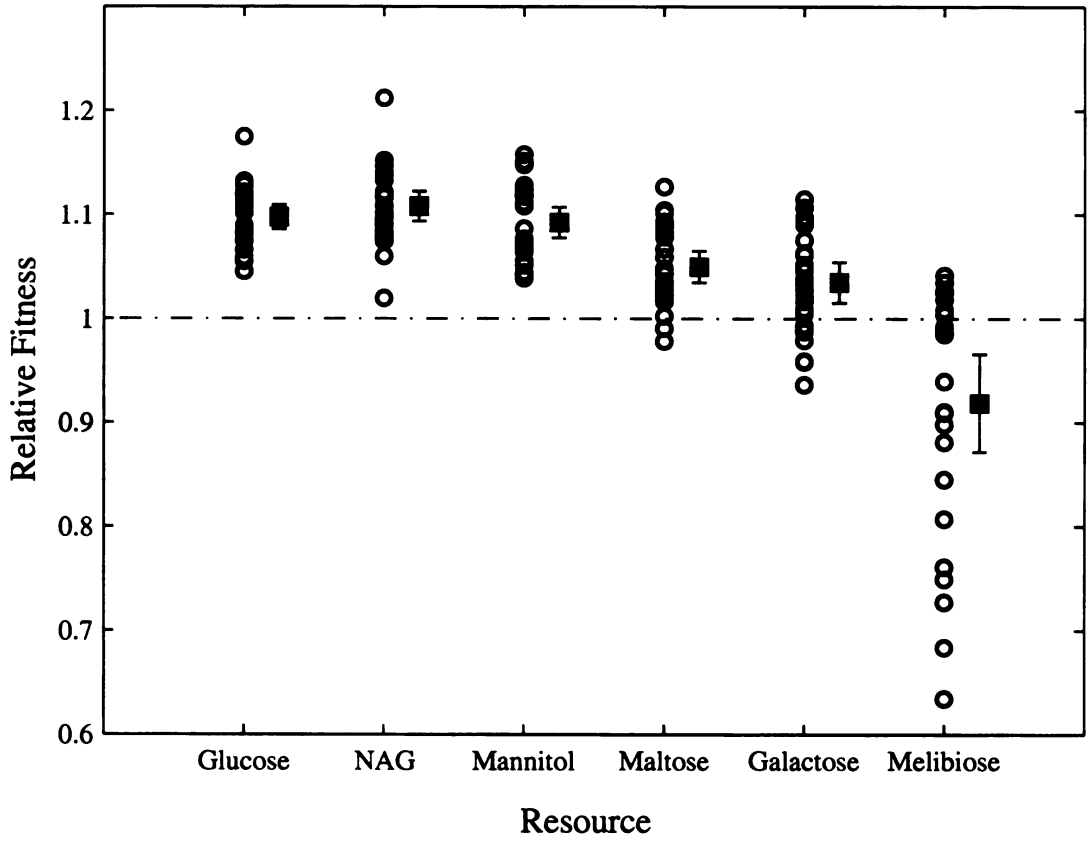


Figure 1. Relative fitness of the 27 mutants in glucose and the five novel resources. Each circle represents the mean fitness of a mutant based on three independent measures (except in glucose, where each value is the mean of 15 independent measures.) Squares represent the grand mean fitness in each resource, and error bars represent 95% confidence intervals based on n = 27.

words, these mutants have an average fitness effect of ~10% in glucose, and a typical mutant differs from that average by roughly 3%.

Table 1. Analysis of variance for fitness effects in glucose. Values are based on 15 independent estimates for each genotype, with three missing values in total. Block and genotype are random effects.

Source	df	MS	*F*	*P*
Genotype	26	0.01273	5.48	< 0.0001
Block	14	0.01245	5.36	< 0.0001
Error	360	0.00232		

Do Beneficial Mutations Have Pleiotropic Effects in Other Resources?

Next we sought to determine whether mutations that are beneficial in glucose have fitness effects in other resources by competing them against their ancestor in five novel carbon sources. In these competitions, a relative fitness of one means that the genotype in question has a fitness equal to that of the ancestor, and thus there is no pleiotropy in such cases. The results of these competitions reveal a significant deviation from neutrality in all five novel resources for the mutants as a group (see confidence intervals, Fig. 1). In addition, the predominant fitness effect in these alternative environments is positive: in 4 of 5 resources, the grand mean fitness was significantly greater than one (see Fig. 1). Only in melibiose was fitness reduced on average, with a mean value of 0.92. We also examined the fitness effects of the mutants individually, and we summarize these results in Table 2. Using a two-tailed t-test, we determined which mutants differed significantly in fitness from an ancestral value of 1.0 in each of the five novel resources. (For

comparison, we also present the results for glucose.) The results confirm the findings based on the mutants as a group: where fitness effects differed significantly from the ancestor, the form of that pleiotropy was predominantly positive. Although 18 of 27 mutants had estimated fitness less than that of the ancestor in melibiose, only 3 of them were statistically significant. If we instead examine the extent of pleiotropy on a per mutant basis, we see that 22 of 27 mutants differed significantly from neutrality in at least one resource, and roughly half differed from the ancestor in two or more resources (Fig. 2). Thus, pleiotropy was common among these beneficial mutations, and its form was overwhelmingly positive.

Table 2. Categorization of the mutant fitness effects by resource. Mean relative fitness was assessed based on three independent estimates of fitness in each resource, except glucose, which was based on 15 estimates. Mutants were determined to be beneficial or deleterious if their fitness differed significantly from 1.0 based on a two-tailed t-test. Abbreviations: GLU, glucose; NAG, N-acetylglucosamine; MAN, mannitol; MAL, maltose; GAL, galactose; MEL, melibiose.

Criterion:	GLU	NAG	MAN	MAL	GAL	MEL
Mean relative fitness ≥ 1	27	27	27	25	20	9
Number significantly beneficial	27	14	12	8	5	0
Number significantly deleterious	0	0	0	0	0	3

Using a two-way analysis of variance, we sought to address simultaneously the contribution of both genotype and resource to the fitness effects seen in novel resources. The results revealed a highly significant genotype × resource interaction (Table 3, $P < 0.0001$). However, the data also indicated a significant departure from the assumption of homogeneity of variances (Levene's test, $F = 2.69$, df = 133, $P < 0.0001$). This effect

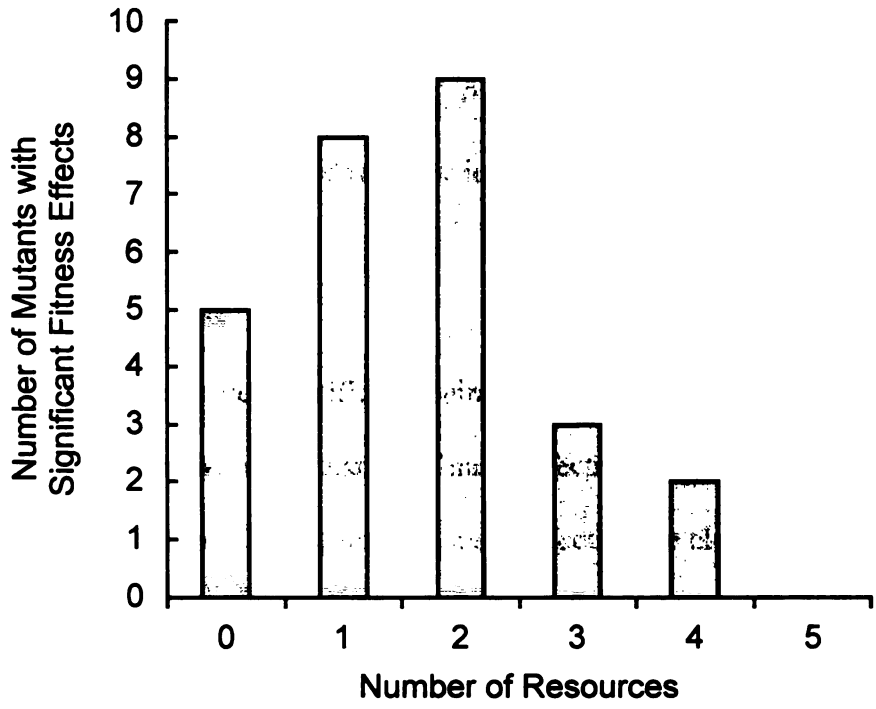


Figure 2. Histogram showing the number of novel resources (out of five) in which mutants had significant fitness effects.

was driven in large part by fitness estimates in melibiose, where the variances both within and among genotypes were much higher than in other resources. To determine whether melibiose was the source of the significant interaction, we re-ran the analysis, but this time excluded the melibiose data. The results show that the interaction was weaker, but still statistically significant ($F_{78, 215} = 1.43$, $P = 0.0237$). The presence of a significant interaction between genotype and resource indicates that the mutants varied in their responses to different resources.

The ANOVA also indicated significant main effects of resource and genotype (both $P < 0.0001$; Table 3), but the interpretation of main effects is problematic when the interaction is significant. To examine the main effects more closely, we performed sets of one-way ANOVAs to examine the effect of genotype within each resource, and the effect of resource for each genotype. Results are summarized in the Appendix, Tables A1 and A2, respectively. These analyses confirm that both genotype and resource were important. The effect of genotype was significant in four of the five novel resources, and marginally significant in the fifth. The effect of resource was significant for 18 of the 27 mutants tested.

Table 3. Two-way analysis of variance for fitness of mutants in the five novel resources. Genotype and resource were treated as random and fixed effects, respectively.

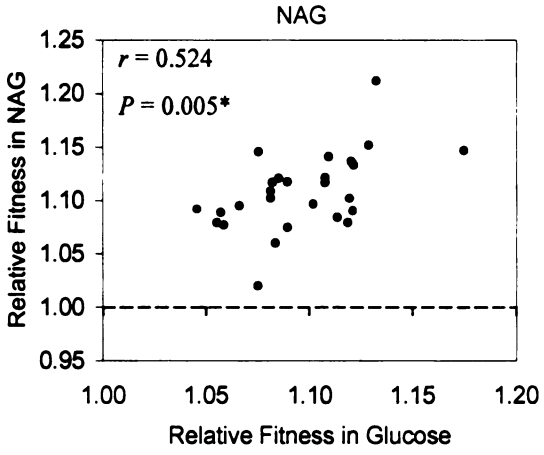
Source	df	MS	*F*	*P*
Genotype	26	0.08427	6.46	< 0.0001
Resource	4	1.68592	51.90	< 0.0001
Genotype × Resource	104	0.03250	2.49	< 0.0001
Error	269	0.01305		

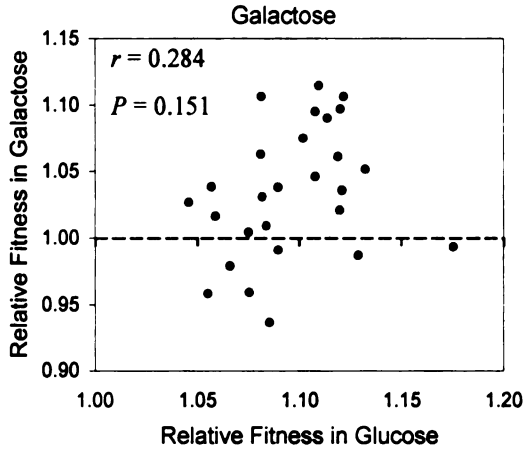
Do Pleiotropic Effects Scale with the Primary Effect?

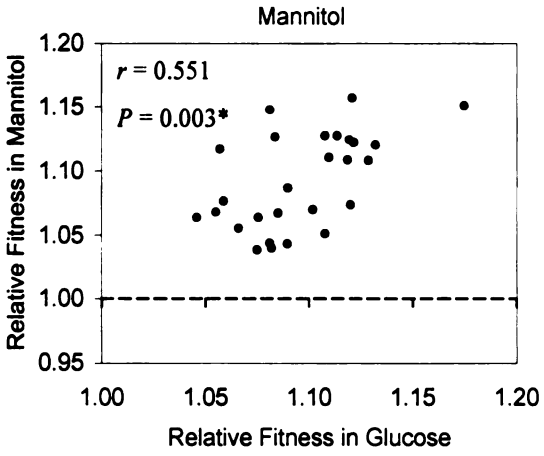
To address whether the pleiotropic fitness effects scale with the magnitude of the direct effects, we calculated the correlation between fitness in glucose and each of the novel resources. For three resources (NAG, mannitol, and maltose), this correlation was highly significant and positive, even after correction for multiple comparisons (Fig. 3). That is, larger benefits in glucose were associated with larger benefits in these novel resources. For the two other resources, galactose and melibiose, no significant correlation was detected. The melibiose data were significantly non-normal (Shapiro-Wilkes $W = 0.969$, $P = 0.003$), so we also calculated a nonparametric Spearman's rank correlation for those data. This test was also nonsignificant ($r_s = 0.092$, $df = 25$, $P = 0.650$). Finally, to assess the robustness of our results, we performed a bootstrap analysis by resampling with replacement from our 27 mutants and recalculating the various correlation coefficients 1,000 times. With a resample size of 27 mutants, this method produced similar results to our parametric tests. The significant results remained so even when our resampled population was reduced to as few as 10 mutants, indicating that our results were robust even to small samples (data not shown).

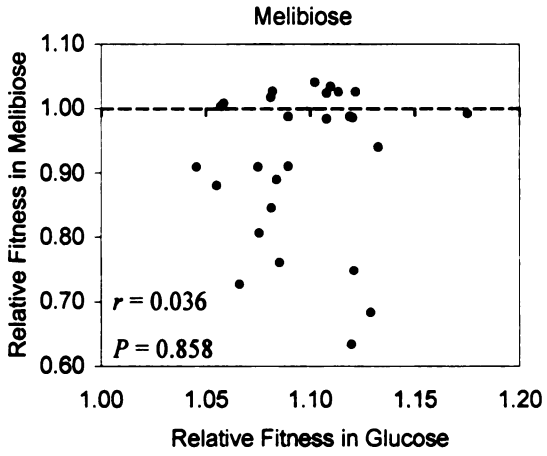
It is interesting to note that the two resources that failed to show statistically significant correlations were the same resources in which average fitness values were lowest. One possibility for the failure to observe a significant correlation is that the mutants with deleterious fitness effects were obscuring a correlation among mutants with beneficial effects, and vice versa, such that positive and negative correlations in different subsets of

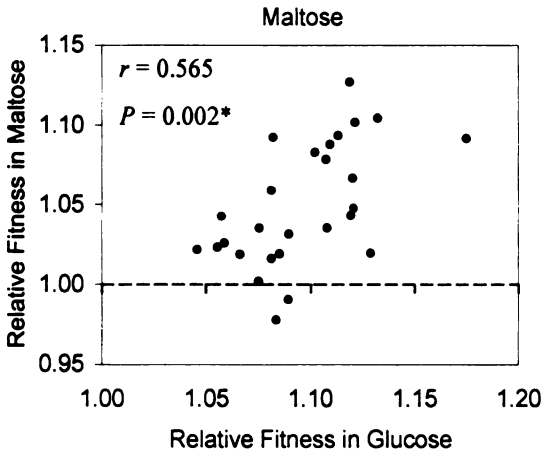
Figure 3. Relationship between fitness in glucose and in five novel resources. Correlation coefficients and their statistical significance are shown on each graph. *Significant following sequential Bonferroni correction.


NAG
r = 0.524
P = 0.005*
Relative Fitness in NAG
Relative Fitness in Glucose
1.25
1.20
1.15
1.10
1.05
1.00
0.95
1.00
1.05
1.10
1.15
1.20


Galactose
r = 0.284
P = 0.151
Relative Fitness in Galactose
Relative Fitness in Glucose
1.15
1.10
1.05
1.00
0.95
0.90
1.00
1.05
1.10
1.15
1.20


Mannitol
r = 0.551
P = 0.003*
Relative Fitness in Mannitol
Relative Fitness in Glucose
1.20
1.15
1.10
1.05
1.00
0.95
1.00
1.05
1.10
1.15
1.20


Melibiose
r = 0.036
P = 0.858
Relative Fitness in Melibiose
Relative Fitness in Glucose
1.10
1.00
0.90
0.80
0.70
0.60
1.00
1.05
1.10
1.15
1.20


Maltose
r = 0.565
P = 0.002*
Relative Fitness in Maltose
Relative Fitness in Glucose
1.15
1.10
1.05
1.00
0.95
1.00
1.05
1.10
1.15
1.20

the mutants were canceling each other out. To address this possibility, we divided the data into two groups: mutants with mean fitness greater than one and mutants with mean fitness less than one. We then calculated separate correlations for these two groups in melibiose and galactose. This analysis allows us to ask whether there is a general trend for mutations with positive pleiotropic effects to scale, and for those with negative effects to scale inversely, regardless of the statistical significance of the individual mutations.

In galactose, the correlation among mutants with fitness greater than one was positive and significant ($r = 0.459$, df = 18, $P = 0.042$), but the correlation for mutants with fitness less than one was also positive and not significant ($r = 0.590$, df = 5, $P = 0.164$). In melibiose, the pattern was similar—removing deleterious mutants resulted in a significant positive correlation ($r = 0.767$, df = 7, $P = 0.016$), but removing mutants with beneficial effects resulted in a correlation that was weakly positive and nonsignificant ($r = 0.111$, df = 16, $P = 0.662$). These results suggest that the deleterious effects of some mutants obscured a positive correlation among mutations with beneficial effects. However, the opposite pattern was not observed: that is, mutants with deleterious pleiotropic effects were not inversely correlated with their effects in glucose. Although the sample size of deleterious pleiotropic mutants was small, even the sign of correlation was not consistent with our hypothesis.

The positive correlations indicate the existence of a predictable, quantitative relationship between the magnitudes of a mutation's direct and pleiotropic effects in three resources. They do not, however, tell us how these effects scale—that is, whether pleiotropic effects

tend to be smaller or larger than their corresponding primary effects. To address the issue of scaling, we calculated the slope of the regression of fitness in each novel resource on fitness in glucose (Table 4). Because our independent variable (fitness in glucose) was measured with error, we performed a Model II Reduced Major Axis regression. The slope of this regression is mathematically equivalent to that of a simple linear regression, scaled by the correlation coefficient of the two variables. The analysis is therefore only appropriate for those comparisons for which there exists a statistically significant correlation—here, for NAG, mannitol, and maltose. Table 4 shows that all three slopes were similar in magnitude and did not differ significantly from a slope of one. Thus, the positive pleiotropic effects were usually similar in magnitude to their direct effects.

Table 4. Estimated Model II regression coefficients (b_{RMA}) for fitness in novel resources versus fitness in glucose and their 95% confidence intervals. A regression coefficient was calculated only if there was a significant underlying correlation.

		95% Confidence Limits	
Resource	b_{RMA}	Lower	Upper
NAG	1.247	0.809	1.685
Mannitol	1.278	0.838	1.717
Maltose	1.315	0.868	1.762

Discussion

Micromutational models of adaptation rest on assumptions about the prevalence, magnitude, and form of pleiotropic effects associated with beneficial mutations, but there is little direct evidence on which to base these assumptions. Here we examine the prevalence, magnitude, and form of pleiotropic effects associated with a sample of

spontaneous beneficial mutations in *E. coli*. Specifically, we ask whether the pleiotropic effects tend to be beneficial or deleterious, and whether they scale in proportion to the primary effects of the mutations. To do so, we examined a collection of mutants, each of which bears a mutation that confers a benefit in glucose. We determined the primary and pleiotropic effects of each mutation by measuring each mutant genotype's fitness in its selective glucose environment and five novel resource environments.

We find that pleiotropic effects of these mutations were common, with the majority having a significant fitness effect in two or more resources. The predominant form of this pleiotropy was positive. We did detect some antagonistic pleiotropy, but it was primarily limited to a single resource, melibiose, and the fitness effects were significant for only 3 of the 27 mutants tested. These antagonistic effects were at times very pronounced, however, with some mutants suffering fitness reductions of more than 30% relative to the ancestor in this resource.

In those resources where pleiotropic effects were positive, the pleiotropic effects of mutations were similar in magnitude to their direct effects, resulting in statistically significant correlations and slopes that did not differ significantly from one. By contrast, deleterious pleiotropic effects did not show a significant inverse correlation with the direct effect. However, the number of these mutations was small, which limited our statistical power. Individual deleterious pleiotropic effects—when they existed—were much larger than their direct effects; the three mutants that were significantly deleterious in melibiose had fitness detriments that were, on average, more than twice their

corresponding benefits in glucose. These few examples provide some weak support for the hypothesis that deleterious pleiotropic effects will be disproportionate to their direct beneficial effect. Nevertheless, there were many large beneficial mutations that lacked deleterious pleiotropic effects in any of the resources tested.

One might suggest that the correlated responses reflect hitchhiking by mutations at other genomic sites rather than pleiotropic effects of the beneficial mutations per se. However, this possibility is unlikely for several reasons, two of which are most powerful. First, Lenski et al. (2003b) sequenced 36 randomly chosen gene regions in bacteria sampled from all 12 of the long-term lines after 20,000 generations. From these data, they estimated a mutation rate for the ancestor of 1.44×10^{-10} per bp per cell generation. Given that the genome contains 4.64×10^{6} bp, this rate corresponds to a genomic rate of 6.68×10^{-4} mutations per cell generation. For the purpose of illustration, let us assume that (i) almost every mutation is selectively neutral in the glucose environment, and (ii) the selective sweep whereby each beneficial mutant arose led to an effective population size of one for the duration. Even under these extreme assumptions, which maximize the potential for hitchhiking, the 400-generation experiments used to select the clones bearing beneficial mutations would have allowed a secondary mutation to hitchhike in only roughly a quarter ($\approx 400 \times 6.68 \times 10^{-4}$) of the cases. The real proportion was presumably much lower. Hence, it is unlikely that more than a handful of the clones carrying the beneficial mutations have any secondary mutations in their genomes. Second, the vast majority of random mutations—including these hypothetical hitchhikers—should be deleterious, not beneficial, when tested in environments that

expose their phenotypic effects. However, the preponderance of correlated responses that we observed in the alternative resource environments were positive, not negative. These data are thus incompatible with the hypothesis that secondary, hitchhiking mutations generated the correlated responses. For these reasons, we are quite confident that the overall patterns of performance we measured on the alternative resources reflect pleiotropic effects of the beneficial mutations selected on glucose.

Our choice of resources was motivated by an earlier study of populations that evolved for 2,000 generations on glucose, many of which showed reduced fitness on several of the resources used here (Travisano and Lenski 1996). In that study, 5 of 12 populations had reduced fitness on maltose, 11 of 12 had reduced fitness on galactose, and all 12 populations showed fitness reductions on melibiose. Our results show a similar qualitative pattern, in that fewer deleterious pleiotropic effects were detected in maltose than in galactose, and fewer in galactose than melibiose. However, our results fail to show that deleterious pleiotropy is common among beneficial mutations, at least with respect to resource use. Given this finding, there are two explanations for the observation that fitness declines were more common over the long-term. One possibility is the mutations were not representative of those occurring over longer times periods. We can speculate that there may be differences between mutations that arise earlier versus later (with the latter possibly having smaller or more targeted effects), but there is no obvious reason to think that the distribution of deleterious pleiotropic effects would also differ. One possibility is that later mutations are more likely to be compensatory and, thus, might be more likely to have deleterious pleiotropic effects. However, mutations

identified to date in the long-term populations appear to be beneficial even on the ancestral background, and therefore compensatory mutations are unlikely to explain our results (Cooper et al. 2003; Cooper et al. 2001; Crozat et al. 2005). A more parsimonious explanation is that our sample was representative of the possible mutations and, in fact, only a minority has deleterious pleiotropic effects on fitness in novel resources. This result implies that the parallel correlated declines in the long-term populations occurred simply because there is a higher probability of having sampled one or more of the subset of beneficial mutations with deleterious pleiotropic effects with increasing time. We might have expected that a similar relationship would hold for resources where fitness effects were primarily positive—that is, that only a small subset of mutations would be responsible for the positive correlated responses. Contrary to this expectation, however, pleiotropic effects were common in these resources and uniform in their direction.

One explanation for the predominance of positive pleiotropy is that the mutations confer increased fitness to aspects of the environment that are in common to all resource regimes—for instance, incubator temperature, daily transfer cycles, and the like. Although these mutations may confer such general benefits, the significant interaction of genotype and resource—as well as the paucity of mutations that were beneficial in melibiose—indicates that at least some of these mutations have resource-specific effects. A second (and not mutually exclusive) possibility is that the mutations enhance competitive ability for glucose, but that glucose is functionally similar to some of the other resources. This possibility was also suggested by Travisano and Lenski (1996), who noted that similarity in performance on alternative resources following long-term

evolution in glucose was highest for those resources that shared their mechanisms of uptake with those used for glucose. Their explanation is supported here by the fact that pleiotropy was most often positive for mannitol and NAG, which share with glucose the phosphotransferase system (PTS) for transport across the inner membrane of *E. coli*. Melibiose and galactose do not involve the PTS, which may explain why fitness effects were generally lower in these resources. The high fitness values observed in maltose, however, are more puzzling in this regard. Maltose is not a PTS sugar and also uses a different porin from that of glucose to diffuse across the outer membrane. Moreover, in a long-term study involving multiple beneficial substitutions, Travisano and Lenski (1996) saw no systematic correlated gains in fitness on maltose, with some populations showing reduced fitness in maltose and others showing fitness increases. Maltose is a glucose dimer, however, and once inside the cell and cleaved, its subsequent catabolism should be identical to that of glucose. Thus, while the pattern of pleiotropic effects seen in NAG and mannitol points to the PTS as a possible target of selection, the finding of positive pleiotropy in maltose suggests that at least some of the mutations included in this study may target later catabolic steps, or otherwise produce similar benefits in glucose and maltose (Travisano 1997). In general, we expect that the greater the similarity in the uptake and catabolism between the novel resource and glucose, the greater the probability that mutations will have similar fitness effects in the two resources.

Despite differences in the fitness effects of mutations in PTS versus non-PTS resources, there remains substantial heterogeneity in the correlated responses, even when we limit our consideration to non-PTS resources only. For example, galactose and melibiose are

both non-PTS resources, and yet beneficial mutations were, on average, positive in galactose and deleterious in melibiose. This result is inconsistent with any single model of how these mutations affect fitness. Travisano and Lenski (1996) reached a similar conclusion, but we note that two additional explanations that could account for the variability in their experiment—drift and epistatic interactions among beneficial mutations—are not in force here. Thus, our study shows that variability in the correlated responses derives solely from the direct effects of different beneficial mutations within the same genetic background.

We emphasize that there are several important caveats to our findings. First, while our results show that positive pleiotropic effects were more common than deleterious pleiotropic effects, our assessment is clearly affected by the choice of resources—it is easy to imagine that a different set of environmental conditions might have produced different results. Moreover, by examining the fitness effects of these mutations on novel resources, we limited the scope of pleiotropy to a set of functionally related traits. In fact, it seems likely that these mutations have effects on other, unmeasured traits, and thus may entail fitness consequences in other environments, including natural ones (Fry 2003; Service and Rose 1985). Pleiotropy can also arise from trade-offs among fitness components within a given resource environment, and this would be an interesting area for further study. A logical set of candidates for this sort of pleiotropy would be performance levels during the distinct series of physiological stages associated with the feast and famine conditions of our serial transfer regime. Following transfer to fresh

medium, there is a period of acclimatization and preparation for growth (lag phase), followed by a period of maximal growth (exponential phase), decelerating growth and, finally, survival after resources have been depleted (stationary phase). Hypothetically, for example, a beneficial mutation might shorten the duration of lag time, thereby speeding the transition to exponential growth, but carry the pleiotropic cost of reduced ability to survive during stationary phase. However, a previous study using populations evolved for 2,000 generations under the same conditions found that these populations were generally improved for several fitness components, with little evidence for reduced performance during any other phase (Vasi et al. 1994).

Finally, an important finding of the current study is the variability in the responses of these mutants to different resources. This observation suggests that the chance occurrence of different mutations in different populations could be an important determinant of their future evolutionary directions (Mani and Clarke 1990; Travisano and Lenski 1996). Moreover, the variation among mutations in the form and extent of their pleiotropy raises interesting questions about the mechanistic bases of these effects and, in turn, about the genetic bases of these adaptations. The wealth of knowledge regarding the physiology and genetics of *E. coli* makes it an ideal system to begin establishing a clear mechanistic link between the action of natural selection on different phenotypes and the underlying genetic changes. Accordingly, we have begun sequencing several candidate genes in each of these mutants. One of these candidates is *spoT*, in which Cooper et al. (2003) found beneficial mutations in 8 of 12 populations that evolved for

20,000 generations under the same conditions and from the same ancestor as the current study. In Chapter 2, we address the molecular basis of these adaptive changes and the extent to which differences in the pattern of their pleiotropic effects may reflect different underlying genetic changes.

CHAPTER 2

THE MOLECULAR BASIS OF PARALLEL AND DIVERGENT PHENOTYPIC RESPONSES IN *ESCHERICHIA COLI*

"...in dealing with such a complex character as selective value, the essential uniqueness of each allele must never be forgotten." –Wright 1968

A major challenge in the study of adaptation is to demonstrate a causal relationship between the action of natural selection on different phenotypes and the underlying genetic changes (Jones 1998; Nachman 2005; Orr and Coyne 1992). This difficulty was aptly described as a "chasm" by Phillips (2005), who emphasized that studies often fall into two broad categories: those that can identify sources of molecular variation, but have limited knowledge of its adaptive significance, and those that can identify ecologically important traits but have limited knowledge of their genetic bases. Among the open

questions regarding the genetic basis of adaptation, an important one concerns the distribution of phenotypic effects associated with different alleles at a given locus (Phillips 2005). Even strict definitions of parallelism implicitly assume that independent substitutions in the same locus are phenotypically equivalent (Schluter et al. 2004). However, as Wright cautioned, multiple alleles at a single locus may differ from one another not just quantitatively, but also qualitatively, owing to their unique pleiotropic effects (Wright 1968, p. 61). While the existence of pleiotropy is hardly in doubt, little is known about the distribution of pleiotropic effects of adaptive mutations and how these effects differ within and across loci.

Previously, we described the phenotypic effects associated with a large sample of spontaneous beneficial mutations that arose from a common ancestor in *Escherichia coli* (Ch. 1). We examined the direct effects of these mutations in the glucose environment in which they evolved, as well as their correlated effects on fitness in five novel resources. Here, we describe the results of work to identify the genetic bases of these adaptations by sequencing candidate loci. By specifically associating our earlier phenotypic measures of divergence with the underlying genetic changes, we assess the extent to which the direct and correlated effects of different substitutions vary within and across loci. We find that substitutions within a locus tend to produce distinct phenotypic clusters. However, this result was driven in part by the large number of mutations that arose in one locus, *spoT*. Other loci showed a more mixed pattern, and in some cases, the precise substitution mattered for the spectrum of direct and pleiotropic effects observed.

Materials and Methods

Collection of Mutants - Isolation of the clones has been described in detail elsewhere (Rozen et al. 2002). Briefly, 30 replicate populations were founded with each of two clones that were isogenic except for a single neutral marker that indicates the ability (Ara^+) or inability (Ara^-) to catabolize arabinose. This difference allows clones of opposite marker states to be distinguished when plated on tetrazolium indicator agar. Populations were founded with Ara^+ and Ara^- clones in the following ratios: 1:100, 1:10, 1:1, 10:1, 100:1. Cultures were propagated daily in a glucose minimal medium, according to the protocol described in Lenski et al. (1991). Cultures were plated periodically to assess the relative proportions of the Ara^+ and Ara^- states; a sustained increase in the frequency of either state indicated that a beneficial mutation had arisen in this subpopulation, at which point clones of both marker states were isolated and saved. Owing to the stochastic occurrence of beneficial mutations, populations were collected at varying timepoints; however, no population evolved beyond 400 generations.

Fitness Assays – All 30 clones were competed against their common ancestor (of the opposite marker state) in six different carbon sources: glucose (the resource on which they evolved), mannitol, maltose, NAG, galactose, and melibiose. The competition assays are described in detail in Chapter 1. Three mutants were found not to have increased fitness relative to the ancestor in glucose, and thus, are unlikely to carry any beneficial mutations. However, to verify this expectation, we included these genotypes in our sequencing

Candidate loci – Five genes or gene regions (*pykF*, *nadR*, *hok/sok*, *spoT*, and an upstream noncoding region of *pbp-rodA*) were identified as candidate loci for adaptation based on previous studies in populations evolved for 20,000 generations (Cooper et al. 2003, Woods et al., manuscript; Cooper et al. 2001; Schneider et al. 2000). The clones used in the current study were derived from the same ancestor and evolved under the same culture conditions as these long-term populations, and thus, these loci were candidates for mutations in these populations as well. A brief description of each gene is provided in Appendix B.

DNA Sequencing – Primers were designed to cover overlapping regions of the *nadR*, *spoT*, *hok/sok*, and *pykF* genes, as well as a region upstream of the *pbp-rodA* locus. PCR products were purified on a GFX™ column prior to sequencing, and all sequencing was done using an ABI automated sequencer. The genes or gene regions of interest were sequenced in their entirety for all 30 clones at least once. All sequences were compared to that of the ancestor and conflicts that could not be resolved by eye were re-sequenced. Candidate mutations were confirmed only if they could be detected on both strands and in sequences arising from a minimum of two independent PCR reactions.

Phenotypic Screening for Rbs⁻ Mutants – Previous studies of populations evolved for 20,000 generations on glucose found that all populations lost the ability to catabolize ribose, owing to a deletion in the ribose (*rbs*) operon (Cooper et al. 2001). To screen for possible ribose deletion mutations, all clones were examined for their ability to grow on ribose. Clones were inoculated into 100 μL of Luria Broth in a 96-well plate and allowed

to grow overnight at 37°C. The following day, 1 μL from each well was transferred into wells containing 100 μL of Davis minimal (DM) medium supplemented with 250 μg/mL of glucose, and allowed to grow and acclimate for an additional 24 hours. On the third day, 1 μL of each culture was transferred to 100 μL of DM supplemented with 250 μg/mL ribose. Each genotype was independently transferred to two wells on the plate, and known *rbs*$^{+}$ and *rbs*$^{-}$ strains were included as positive and negative controls, respectively. Readings of the optical density of each well were taken every 5 minutes for 24 hours using a VersaMax automated microplate reader and used to construct growth curves.

Statistical Analyses

Analysis of Fitness Effects Within and Between Loci – To assess the phenotypic similarity of genotypes with mutations in the same locus, we performed an analysis to cluster the genotypes according to their fitness effects in the six resources. The distance metric used for the analysis was a normalized Euclidean distance. Linkage proceeded by an algorithm that iteratively joined the closest genotype to a given cluster, where the location of the cluster was defined by its centroid. Other methods produced similar results (e.g., Minkowski distance and/or linkage based on average distance.) Clustering was performed in SYSTAT (SYSTAT 2002).

To determine whether there was a significant association between the phenotypic effects of mutations and the underlying genetic changes, we performed an analysis of similarity

(ANOSIM) using the Primer-E software package (Clarke 1993; Clarke and Gorley 2001). The analysis of similarity uses a distance matrix to rank all pairs of genotypes from most similar (lowest rank) to least similar (highest rank), and then calculates the difference in the mean rank of the between-group comparisons to that of the within-group comparisons. The resulting statistic is scaled according to the number of possible combinations to produce an R-statistic that ranges from 0 to 1. In our study, an R-statistic of zero would indicate that the fitness effects of different mutations are as heterogeneous within a locus as they are across loci. A value close to 1 would indicate that mutations in the same gene are more similar to one another than they are to mutations in other genes. The statistical significance of the ANOSIM result is determined by a permutation test where the genes are assigned at random to the genotypes and the analysis is repeated. Where feasible, all possible permutations of the data set were performed; otherwise, 1,000 permutations were performed. The permutations were used to determine the distribution of the R-statistic under the null hypothesis. The observed value was compared to this distribution in order to assign statistical significance.

The ANOSIM analysis allows us to determine whether there is more variation in mutant effects among than within loci. However, it cannot tell us whether there is significant variation in fitness effects between mutations in the same locus. To address this question, as well as to assess the relative contributions of direct versus correlated responses to the overall variance in phenotypic effects, we performed a series of nested ANOVAs on

different subsets of the data. First, using only the genotypes with identified mutations, we determined what proportion of the variance in glucose could be attributed to the locus versus the genotype. This analysis was performed as a nested one-way ANOVA in SAS. To examine the pleiotropic effects of these mutations, we performed a two-way nested ANOVA on the fitness effects of the mutants in the five novel resource environments. Factors included in the model were: resource, locus, genotype(locus), and their respective interactions. Where a sufficient number of mutations were identified within a locus, we also performed separate ANOVAs on each locus alone. All analyses were performed in SAS using PROC GLM (SAS Institute 1999). Genotype and locus were treated as random effects, whereas resource was treated as a fixed effect. F-tests were constructed in a manner analogous to that described in Goldberg and Scheiner (2001) for a two-way nested ANOVA, except that a Satterthwaite correction was used where the data was unbalanced. For simplicity we round the degrees of freedom to the nearest whole number when we present such F-tests in the text. Finally, to assess how the direct and pleiotropic effects contributed to the clustering of different genotypes, we performed a principal components analysis. The data used in the analysis were the relative fitnesses of the mutants on six resources: glucose, NAG, mannitol, maltose, and galactose. All variables were standardized to z-scores prior to analysis to achieve similar means and variances. The analysis was performed using SYSTAT (2002).

Results

Sequencing of candidate loci

Sequencing of the five genes or gene regions uncovered a total of 21 mutations: 13 in *spoT*, 5 in *nadR*, and 1 each in *pbp-rodA*, *pykF* and *hok/sok*. In one case, two mutations were found in the same clone (*spoT* and *pykF* in 9990). All of the clones were able to grow on ribose, indicating that their ribose operons were intact. A list of identified mutations is provided in Appendix C, and the locations of the mutations found in *spoT* and *nadR* are shown relative to those identified previously in long-term populations in Appendix D. As expected, no mutations were found in the three clones that previously failed to demonstrate a significant fitness increase relative to the ancestor. Owing to their lack of a beneficial mutation, they have been excluded from the remaining analyses. In a further three cases, sequencing uncovered the same mutation in independently isolated clones. The analyses presented here include these duplicates, as they provide a useful reference point to which the phenotypic similarity of non-identical mutants can be compared. However, to ensure that they did not influence our results, we repeated all of our analyses on reduced data sets where we excluded one member of each duplicate pair. Exclusion of these genotypes did not alter any of our conclusions.

Concordance between measures of genotypic and phenotypic similarity

To examine the association between the phenotypic effects of these mutations and their underlying genetic changes, we used a hierarchical cluster analysis to group the mutants

according to their fitness effects on the six different resources. This analysis allows us to assess the phenotypic similarity of the different mutants. To assess the concordance between this measure of phenotypic similarity and the underlying genetic changes, we then overlaid the diagram with the identity of the locus where mutations were subsequently identified. The resulting diagram is presented in Figure 1, with highlighting used to delineate major clusters. If no mutations were found in any of the sequenced genes or gene regions, the genotype has been left unlabeled.

Figure 1 shows that mutations within a locus tend to cluster together, indicating that the fitness effects within a locus tended to be more similar to one another than the fitness effects across loci. However, it is also important to note that many clusters include genotypes for which no mutations were identified in any of the sequenced loci. Because these genotypes are certain to carry some beneficial mutation (owing to their increased fitness on glucose), all of the clusters necessarily encompass more than one locus. Finally, genotypes with the exact same substitution map to the same clusters, but in no case are they most similar to each another. This result indicates that there is likely little differentiation in the phenotypic effects among mutations within a locus. We address this possibility more directly in later sections.

There are also several interesting exceptions to the general pattern of clustering. For instance, two of the identified *spoT* mutations are highly divergent from all other mutants with substitutions in *spoT* (Figure 1). One of these genotypes was found to carry a secondary mutation in the *pykF* locus; its divergence is therefore not surprising. The

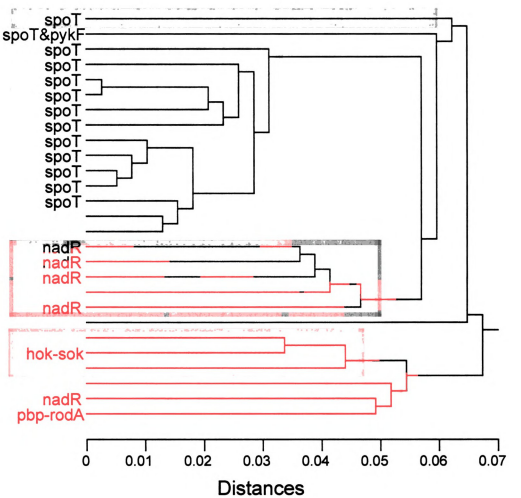


Figure 1. Cluster diagram produced by a hierarchical cluster analysis, based on estimates of relative fitness of the genotypes on six different resources. The distance metric used was a normalized Euclidean distance. The linkage algorithm iteratively joined objects with the shortest distance to an existing cluster, where the cluster was determined based the location of its centroid. Other cluster algorithms produced similar results. Overlaid on the diagram are the identities of the loci where mutations were subsequently found. Genotypes where no mutations were found in the six different loci have been left unlabeled. Coloring has been added to the diagram to highlight the major clusters. Symbols represent pairs of genotypes that had the same mutation (see Appendix C). Images in this thesis are presented in color.

other genotype (highlighted in gray; Figure 1) is also interesting, as it is the only mutation that arose in the region of overlap between the ppGpp synthetase and ppGpp hydrolase regions of the gene (see Appendix B), although the boundaries of these regions are not known precisely (Gentry and Cashel 1996). One possibility is that the fitness effects of mutations in this region differ from those in other regions of the gene.

To assess the statistical support for the clustering pattern observed in Figure 1, we performed an analysis of similarity (ANOSIM), which is summarized in Table 1. The ANOSIM tests whether genotypes within a locus are more similar to one another phenotypically than they are to genotypes in different loci. The ANOSIM is similar to a nonparametric ANOVA, except it uses a permutation test to assign statistical significance. The global test shows an R-statistic equal to 0.8, which is highly significant ($P = 0.001$; Table 1), and substantially outside the range of all other estimates based on random permutations (Figure 2). This result in agreement with what can be observed in Figure 1, indicating that the fitness effects within a locus are more similar to one another than expected by chance. In addition, the ANOSIM on individual pairs of loci confirm that the phenotypic effects of mutations in *nadR* and *spoT* also differ significantly from one another. Other comparisons were not significant, but the number of these mutations was very small. In all comparisons involving the *spoT* locus, the observed value of the R-statistic was the most extreme of all possible permutations of the data (Table 1). This result indicates that *spoT* is highly divergent in its fitness effects compared to other loci, a result that is consistent with the cluster diagram in Figure 1.

Table 1. Results of the analysis of similarity (ANOSIM)

Groups	R-statistic	Significance level	Number of Permutations: Possible	Performed	>= to observed R
Global Test:					
	0.8	0.001	7,054,320	1,000	0
Pairwise Tests:					
spoT, nadR	0.821	0.001	6188	1,000	0
spoT, spoT&pykF	0.833	0.077	13	13	1
spoT, hok-sok	0.902	0.077	13	13	1
spoT, pbp-rodA	0.942	0.077	13	13	1
nadR, spoT&pykF	0.920	0.167	6	6	1
nadR, hok-sok	-0.160	0.667	6	6	4
nadR, pbp-rodA	0.320	0.333	6	6	2

Table 1. Results of an analysis of similarity (ANOSIM) to determine whether mutations within a locus were more similar to one another in their fitness effects than mutations in different loci. A distance matrix was first constructed using the relative fitness of the mutants on six different carbon sources. The distance matrix was then used to assign a rank to every possible pairwise combination of genotypes, ranging from most similar (lowest rank) to most different (highest rank). R was calculated as the difference in mean rank of genotype pairs with mutations in different loci to those within a locus, scaled according to the number of combinations. To determine the significance of the value, a permutation test was used where loci were assigned at random to the different genotypes and the calculations were then repeated. Where possible, the statistic was calculated on all possible permutations of the data. Otherwise, a random sample of 1,000 permutations was estimated.

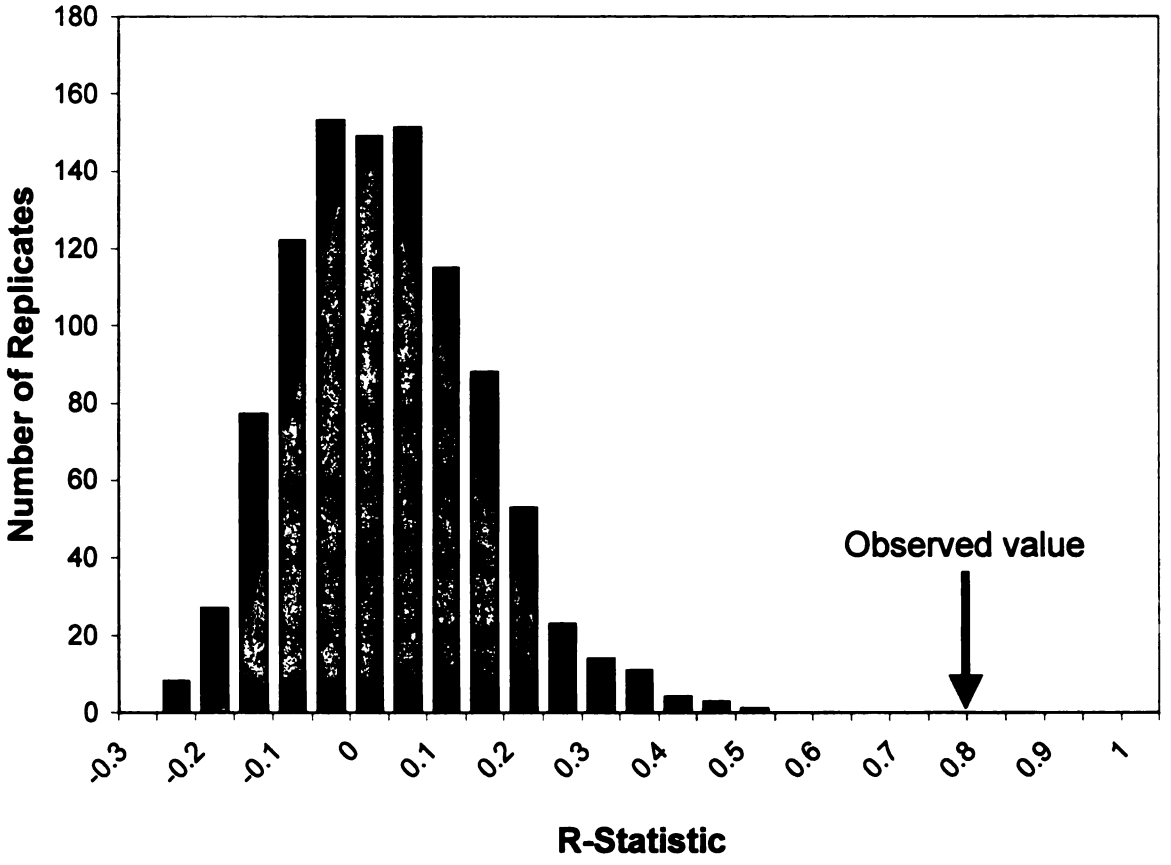


Figure 2. Histogram based on the results of bootstrapping the R-statistic from the global analysis of similarity. The histogram is based on 1,000 replicates where the gene identity was randomly assigned to each genotype.

The hierarchical clustering analysis demonstrates that different mutations in the same gene produce a unique phenotypic signature. However, it provides no information about the contributions of the different resources to this clustering pattern. To develop a clearer understanding of how the different resources contribute to phenotypic clustering, we also performed a principal components analysis (PCA) on the fitness effects in the six resources. The results of this analysis showed that the first and second principal components explained 46.0% and 24.3% of the variance, respectively.

The factor loadings plot is shown in Figure 3. Interestingly, the vectors for the three resources (glucose, NAG, and mannitol) that use a common uptake mechanism (the phosphotransferase system, or PTS) are virtually indistinguishable in their contributions to the first two principal components. However, the three non-PTS resources (maltose, galactose, and melibiose) are divergent from both the PTS resources and from each other. This result is interesting in light of previous work suggesting that the PTS system is the target of selection in a glucose-limited environment (Travisano and Lenski 1996). Moreover, previous work on these mutations identified strong positive correlations between fitness in glucose and fitness in NAG, mannitol, and maltose (Chapter 1). The strong positive correlation between fitness in glucose and maltose was puzzling in light of its classification as a non-PTS resource. The principal components analysis thus suggests that there is underlying heterogeneity in the response of these mutants on maltose compared to the PTS resources, something that was not evident previously based solely upon the individual correlations. Finally, Figure 4 shows a plot of the factor scores for each mutant, with the locus indicated where it has been identified. The results from

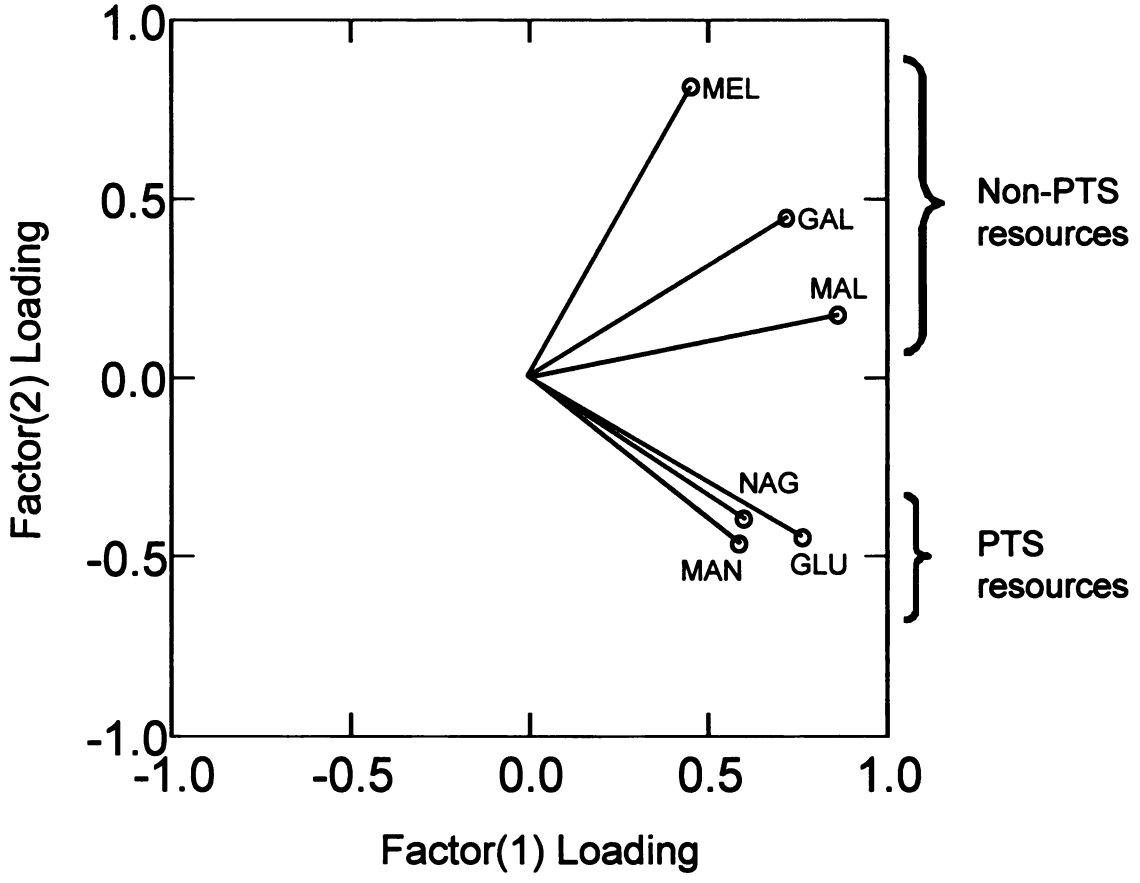


Figure 3. Factor loadings plot for the principal components analysis. Variables included in the analysis were relative fitness in six resources: glucose, NAG, mannitol, maltose, galactose, and melibiose. All variables were standardized prior to the analysis in order to normalize the variances. PTS resources share a common mechanism for transport across the inner membrane of the cell, whereas non-PTS resources use a variety of different mechanisms. See text for greater detail.

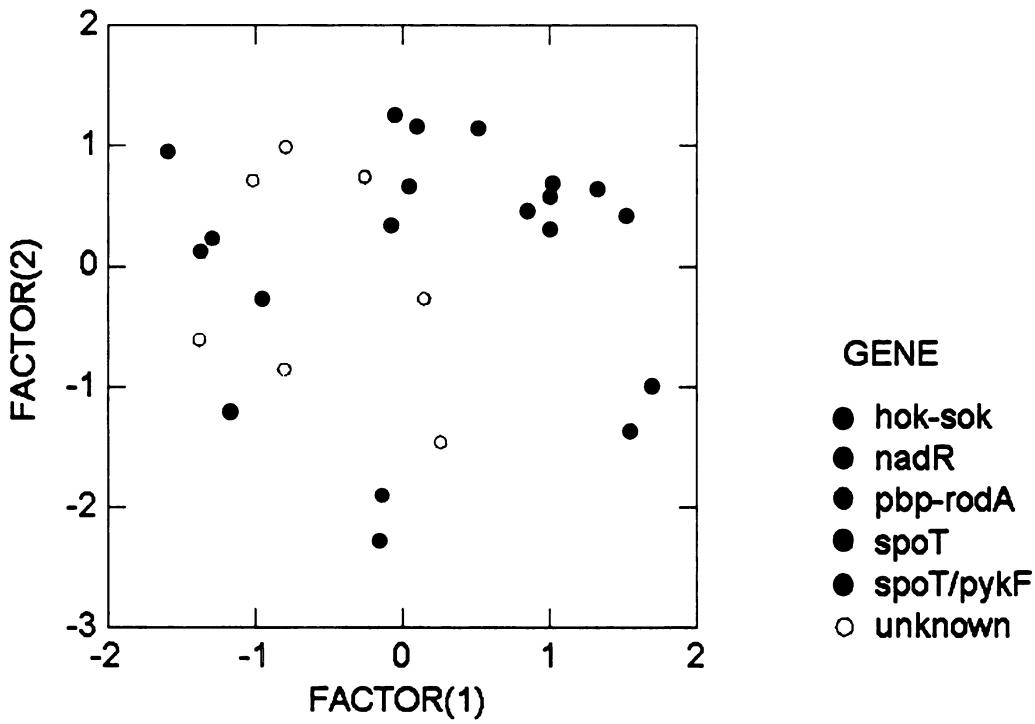


Figure 4. Plot of the genotypes according to their factor scores for the first two principal components. The identity of the loci where mutations were found has been overlaid onto the plot. Images in this thesis are presented in color.

the principal components analysis are similar to those obtained from the hierarchical cluster analysis (Figure 1), showing that mutations within a locus tend to cluster together.

Direct versus pleiotropic effects of mutations

The results presented above show that the fitness effects on six carbon sources are sufficient to distinguish the different loci. However, it is not clear whether these effects are more strongly driven by differences in direct effects (that is, fitness effects in glucose), or differences in their pleiotropic effects on other resources. To examine this issue more closely, we performed a series of ANOVAs on the fitness effects in glucose, the selective environment, and the five novel resources. In each case, we performed a large nested ANOVA that included all the genotypes for which a mutation had been identified. However, because the five loci were differentially represented, the ANOVAs were unbalanced. Thus, we performed additional ANOVAs to address heterogeneity in fitness effects in *nadR* or *spoT* only; other loci were only represented once, and could not be examined individually. We also excluded the genotype with the *spoT-pykF* double mutation from our analysis because we cannot disentangle the fitness effects of the two mutations.

To examine the direct effects of all these mutations, we performed a one-way nested ANOVA on the fitness effects in glucose. The results of this analysis are summarized in Table 2, and show that there was no overall effect of genotype ($P = 0.201$) and only a marginally significant effect of locus ($P = 0.100$). However, these results differ from what is obtained when we perform separate ANOVAs on *spoT* and *nadR* mutants in

glucose, which show a significant effect of genotype within *nadR* ($F_{4,69} = 3.82$, $P = 0.007$), but not within *spoT* ($F_{11,167} = 1.28$, $P = 0.239$). This result indicates that there is heterogeneity in the direct effects of different substitutions within the *nadR* locus, but that different *spoT* mutations are not phenotypically distinct from one another, at least with regard to the carbon sources we examine here.

In contrast to the limited variation observed in the direct glucose effects, substantial variation is evident when we examine correlated responses in the five non-selected resources. These results are summarized in Table 3, and show a significant effect of locus ($P = 0.036$), resource ($P = 0.001$), resource × genotype(locus) ($P = 0.025$), and resource × locus ($P < 0.0001$). Only the effect of genotype was nonsignificant ($P = 0.540$). These results suggest that whereas there is relatively limited variation among loci in their direct effects, they differ substantially in their pleiotropic effects. Moreover, loci vary in their pleiotropic effects depending on the resource (i.e., there is a significant resource × locus interaction), as do different substitutions within a locus (significant resource × genotype(locus) interaction). With regard to the possibility of differences among loci in their pleiotropic effects, the results of this analysis are similar to those of the ANOSIM analysis. However, the significant resource × genotype(locus) interaction is new, indicating that even within a locus, different substitutions can entail distinct pleiotropic effects.

Table 3. Analysis of variance on fitness effects in five novel resources

Source	df	MS	F	*P*
Resource	4	0.3336	10.54	0.0006
Locus	3	0.1402	4.11	0.036
Genotype(locus)	15	0.0029	0.92	0.540
Resource × locus	12	0.0358	7.67	<0.0001
Resource × genotype(locus)	60	0.0047	1.48	0.0253
Error	189	0.0032		

Table 3. Results of a two-way nested ANOVA examining the fitness effects of the mutants on 5 novel resources. A Satterthwaite approximation was used to determine the denominator of certain F-tests owing to imbalance in the data set; however, the statistical significance of tests were not affected by this correction.

Finally, separate analyses on the *spoT* and *nadR* mutants shed further light on the effect of mutations within a locus. When fitness effects of *spoT* mutants in the five novel resources are examined, there is a significant effect of resource ($F_{4,44} = 24.10$, $P < 0.0001$), but not of genotype ($F_{11,44} = 1.35$, $P = 0.232$) or genotype × resource ($F_{44,119} = 1.01$, $P = 0.467$). This result indicates that the mean fitness of *spoT* mutants varies according to the particular resource in question, but that different *spoT* mutants respond to these resources in a correlated fashion. The *nadR* mutants showed a similar pattern, with no significant effect of genotype ($F_{4,16} = 0.11$, $P = 0.979$), but a significant effect of resource ($F_{4,16} = 12.19$, $P < 0.0001$). However, in this case genotype × resource ($F_{16,50} = 3.64$, $P = 0.0002$) was significant, indicating that different *nadR* mutations entail different pleiotropic effects. Overall these results indicate that whereas different mutations in the *spoT* locus are phenotypically equivalent to one another, the effects of mutations in the *nadR* locus are highly dependent on the precise substitution.

Discussion

We quantified the extent of parallel and divergent phenotypic responses that result from the substitution of different beneficial mutations in *Escherichia coli*. A key strength of our experimental design was the ability to assess the phenotypic effects of individual mutations, and thus to determine how much of the variation in the effects of different genotypes is subsumed at the level of the gene. Our results show convincingly that mutations within a locus are more similar to each other than they are to mutations in other loci. This result, however, was largely driven by a single locus, *spoT*. Although some variation was detected in the direct effects of mutations within and across genes, far more variation was detected when we examined their pleiotropic effects on fitness in alternative resources. This result suggests that substitutions that appear similar in one respect can nevertheless entail different pleiotropic effects, and thus, different consequences for the direction of subsequent evolution.

Despite having identified the genetic basis of many of these adaptations, we know surprisingly little about why these changes are beneficial. Nevertheless, an examination of these mutations on a variety of resources provides us with some useful information, and allows us to outline a few general trends. First, *spoT* mutations tend to be beneficial across all resources tested, suggesting that their benefit may result from a general mechanism. One of the explanations suggested previously for the advantage provided by *spoT* mutations would be consistent with a general benefit (Cooper et al. 2003). These authors noted that a *spoT* mutation was sufficient to cause a shift in the pattern of stable

RNA regulation within a cell, a shift that has been associated with an increased growth rate in glucose (Sarubbi et al. 1988). This shift is also thought to be part of the response to carbon starvation more generally, and thus is a good candidate mechanism to explain our results. We note, however, that the ANOVA on the *spoT* mutants revealed a strong effect of resource, indicating that the while the mutations were beneficial across all resources, the magnitude of this benefit differed depending on the specific resource. Nevertheless, a general mechanism does not necessitate a general response, particularly for traits that involve the interaction of many component functions, such as resource uptake and catabolism.

The results of this study also showed that *spoT* mutations were far more common than mutations in other loci. There has been considerable debate about the relative importance of changes to regulatory versus structural genes for adaptation (Barrier et al. 2001; King and Wilson 1975; Purugganan 2000; Remington and Purugganan 2003), and thus it is interesting that the majority of adaptive changes would arise in a global regulatory gene. Even more surprising is that these mutations would have fewer detectable deleterious pleiotropic effects than other mutations. However, our detection of pleiotropy was limited to effects on the use of alternative carbon sources, and thus, it would be interesting to know whether these mutations would prove disadvantageous in a more extreme environment. Because *spoT* is thought to promote entry into stationary phase by increasing transcription of the relevant genes, an obvious place to look for a trade-off is stationary phase survival. However, previous studies of the long-term populations (many of which are known to harbor *spoT* mutations) showed no reduced ability to survive

during prolonged stationary phase (Vasi and Lenski 1999). The possibility remains, however, that some of these mutations interact epistatically and may even compensate for one another—for example, mutations in the *pbpA* locus that prevent cell division have been shown to be rescued by increases in ppGpp, the molecule encoded for by the *spoT* locus (Lutkenhaus and Mukherjee 1996). Thus, caution must taken when attempting to infer the phenotypic effects of single mutations from genotypes that bear multiple substitutions.

A surprising result from this study was that the *nadR* mutations in the short-term lines were predominantly insertions or deletions (3 deletions, one IS insertion, and one nonsynonymous point mutation). In contrast, the *nadR* mutations in the long-term lines were all nonsynonymous point mutations, with the exception of one IS insertion (Woods, unpublished observations). NadR is a bifunctional protein, with the N-terminal portion important for preventing transcription of the NAD biosynthetic genes, and the C-terminal portion important for transport of NMN into the cell (Penfound and Foster 1996, see also Appendix B). One possibility is that deletion mutations arise more readily in *nadR* than nonsynonymous substitutions, but have greater pleiotropic effects, since they are more likely to affect both functions of the gene, particularly when they arise very close to its start, as these do. Thus, there could be selection occurring among different *nadR* mutations. This possibility is possibly supported by the finding of significant variation among *nadR* mutants in both their direct and pleiotropic effects. However, there was no clear pattern to how these fitness effects vary depending on the type of mutation or its

location in the gene. Construction of isogenic strains that vary only in their mutations at this locus will be helpful for addressing this hypothesis more directly.

Previous work on the pleiotropic effects of these mutations also found that many mutants demonstrated fitness reductions on melibiose, three of which were shown to be significantly deleterious on this resource. Two of these three mutations remain unidentified, but the third is a mutation in the *pbp-rodA* locus. Mutations in this locus are known to affect cell size and shape, and thus repeated substitutions in this locus thus may explain the correlated increases in cell size seen in long-term populations. While it is not clear why mutations in the *pbp-rodA* locus would be deleterious in melibiose, differences in PTS and non-PTS resources primarily relate to their mechanisms of uptake from the environment, as their respective catabolic pathways converge soon after transport into the cell (Travisano and Lenski 1996). Thus, mutations that alter cell wall synthesis may be more likely to have differential effects on PTS and non-PTS resources. Interestingly, one of the mutations identified in long-term populations arose in a gene also known to be involved in peptidoglycan synthesis, *glmUS*. Genotypes that cluster with the mutation upstream of the *pbp-rodA* locus (e.g., Figure 1) may thus be good candidates for mutations in this gene.

Interestingly, *pykF* was one of the few loci for which a differential effect in PTS versus non-PTS resources could be predicted *a priori* (Schneider et al. 2000). Pyruvate kinase I, encoded by the *pykF* gene, catalyzes the conversion of phosphoenolpyruvate (PEP) and ADP to pyruvate and ATP. Additionally, PEP is used by the PTS for the phosphorylation

and uptake of sugars across the inner membrane of the cell. Thus, a mutation that knocks out *pykF* may conserve PEP, which can then be used power transport of PTS resources into the cell (Schneider et al. 2000). Thus, *pykF* mutations may provide little or no advantage for the transport of non-PTS resources, and may even be deleterious, depending on whether there is a primary cost associated with the loss of the *pykF* gene. However, because *E. coli* possess a second pyruvate kinase gene, we cannot assume that the loss of this gene would be deleterious. In fact, studies have shown that only one of these genes is required for growth (Ponce et al. 1995). Unfortunately, our study uncovered only a single *pykF* mutation, and it co-occurred with a *spoT* mutation, making it impossible to disentangle their independent effects. However, fitness data for this genotype indicates that it is neutral in melibiose (mean relative fitness = 0.99). Because *spoT* mutations—when they arise alone—appear to be weakly beneficial in melibiose, the *pykF* mutation therefore appears to negate this benefit. This finding thus provides some support for the hypothesis of a cost associated with *pykF* mutations in non-PTS resources.

On a final note, we emphasize that there are many practical benefits to studying the phenotypic effects of single substitutions in naturally occurring variants. First, many studies of bacterial physiology focus on the effects of knock-out mutations, which provide only limited insight into the function of a given gene or the phenotypic effects of a given mutation. By examining the fitness effects of selected mutations—and in particular, their effects under a variety of environmental conditions—we develop a more nuanced understanding of the functions of these genes and the diversity of their pleiotropic effects. More practically, developing a phenotypic signature of different

genetic changes provides a useful diagnostic tool for identifying candidate loci in other clones. Finally, understanding the suite of adaptive changes that has taken place in long-term populations evolved for 20,000 generations (Cooper et al. 2003; Lenski et al. 1991; Schneider et al. 2000) clearly requires that we first understand the phenotypic and fitness consequences of individual mutations. However, an important next step will be to examine the fitness effects of these mutations in combination. Previous work has shown that mutations in the *spoT* locus do not provide a fitness advantage on the genetic background of clones from all evolved populations, pointing to the existence of a phenotypically equivalent mutation that has fixed in some populations (Cooper et al. 2003), it is apparent already that the combined effects of these mutations will differ from the sum of their parts. Moreover, elucidating the extent of parallelism and divergence in these populations requires not only that we understand how these substitutions work in concert, but also how one substitution may influence the likelihood of subsequent substitutions, a process that may have a large impact on the trajectory of parallel or divergent evolutionary changes.

CHAPTER 3

ECOLOGICAL SPECIALIZATION AND ADAPTIVE DECAY IN DIGITAL ORGANISMS

Many theories about the origins and maintenance of biological diversity involve specialization and adaptive decay. Specialization describes the process by which organisms become highly adapted to a narrow range of environmental conditions, and may be associated with adaptive decay—the loss of other traits, functions, or abilities that results in the evolution of narrow niche breath. A tendency toward increased specialization is a defining feature of adaptive radiations, as it forms the underpinnings for niche partitioning and character displacement, which promote diversification (Schluter 2000; Simpson 1953).

The process of specialization can result in the loss of other functions in environments in which they are no longer useful, termed adaptive decay. For example, the transition from

a free-living to a parasitic lifestyle is often thought to involve not only adaptations that enable host exploitation, but also extensive decay of other unused functions, presumably necessary for survival outside the host, with parasites showing reduced or streamlined genomes relative to their free-living relatives (Andersson et al. 1998; Cole et al. 2001; Ochman and Moran 2001; Shigenobu et al. 2000).

Both specialization and adaptive decay have been documented in natural populations, but the underlying genetic mechanism remains unclear. Some have hypothesized that there are trade-offs, such that adaptation to one environment inevitably results in loss of adaptation to others (antagonistic pleiotropy hypothesis). Trade-offs may result from an energetic burden associated with maintaining or expressing unused functions, or because improvements to a selected trait directly interfere with the functioning of an unselected trait. An alternative hypothesis is that the loss of specialized features results from relaxed selection, enabling mutations to accumulate in the portions of the genome that encode unused functions (mutation accumulation hypothesis). Which of these mechanisms predominates is important, insofar as they lead to different expectations as to the frequency of specialization and the types of circumstances that promote it. For instance, mutation accumulation requires that the genes that contribute to increased adaptation in alternative environments be distinct and that the environment be heterogeneous (in time or space), so as to give rise to the periods of relaxed selection that enable mutations to accumulate. Alternatively, antagonistic pleiotropy results in constraints that prevent organisms from being simultaneously adapted to many niches and does not require environmental heterogeneity, although it may be aided by it. Generally speaking, if

antagonistic pleiotropy is common, then the process of specialization will be closely tied to that of adaptive decay.

Despite long-standing interest in these two hypotheses, it has been difficult to distinguish between them. A large body of work on adaptive decay (also called regressive evolution) has focused on cave organisms (Jeffrey 2003; Jernigan 1994; Jones 1992). These organisms often exhibit highly convergent and distinctive phenotypes, characterized by loss of pigmentation and reduced visual systems, but with other sensory structures being highly developed, such as antennae. Mutation accumulation hypotheses suggest that the lack of light in caves resulted in relaxed selection and the accumulation of mutations that eventually led to the losses of pigmentation and visual sensory structures. Alternatively, antagonistic pleiotropy hypotheses posit that adaptation to low light levels led to highly specialized sensory structures and that the losses of other traits were a direct result of this adaptation, possibly because the maintenance of unused functions was costly. For example, Darwin hypothesized that eyes are costly to burrowing rodents because they are prone to infection, and thus that their evolutionary loss may have been aided by natural selection (Darwin 1859). Although the extent to which regressive phenotypes reflect mutation accumulation or antagonistic pleiotropy has been a subject of great debate, a recent study of cave fish demonstrated linkage for QTLs associated with both a regressive (eye size) and a "constructive" (body weight) trait, suggesting that either antagonistic pleiotropy or hitchhiking was responsible (Borowsky and Wilkens 2002). In general, increased knowledge of the genetic basis of traits, as well as their evolutionary

dynamics, is expected to shed light on the processes of mutation accumulation and antagonistic pleiotropy.

Experimental approaches provide an alternative to comparative approaches, allowing direct examination of the process of specialization and adaptive decay. A recent study of evolving populations of *Escherichia coli* examined the consequences of long-term adaptation to a simple environment for the evolution of catabolic niche breadth (Cooper and Lenski 2000). Replicate populations of *E. coli* were propagated for 20,000 generations in a medium containing only a single available carbon source, glucose. While the evolved populations were found to have increased ability to compete for and catabolize glucose relative to their ancestor, they also consistently evolved reduced ability to catabolize other resources. Moreover, the identities of these carbon sources were similar across independently evolved populations. This parallelism suggested that they resulted from antagonistic pleiotropy—that is, that the reduction in diet breadth had traded off with improved ability to use glucose. Populations that evolved elevated mutation rates during the course of the experiment also did not show significantly greater losses, contrary to the expectations under mutation accumulation, and thus further indicating that antagonistic pleiotropy was the primary mechanism.

Here we address the process of specialization in a very different medium—an evolving system comprised of self-replicating computer programs that mutate, compete, and evolve in a computational environment. We examine the digital equivalent of diet breadth—the ability of these organisms to perform complex computations that enable

them to garner energy from their environment. In this system, we not only can observe the pattern of evolution associated with specialization and adaptive decay, but we can also examine in detail the underlying genetic processes—that is, we can identify the specific mutations that result in losses of function and determine their fitness effects. We use this knowledge to distinguish between antagonistic pleiotropy and mutation accumulation by asking whether losses of function were the result of neutral or beneficial mutations. Whereas losses that result from mutations that are neutral in the selective environment constitute examples of mutation accumulation, those that result from mutations that are beneficial in the selective environment constitute examples of antagonistic pleiotropy.

Below, we give a brief introduction to the digital life system, *Avida*. Additional information is provided in Appendix E, including a schematic of a digital organism and a glossary of terms. A more detailed description of the system is available elsewhere (Lenski et al. 2003a; Ofria and Wilke 2004; Wilke and Adami 2002), and documentation is available online (http://devolab.cse.msu.edu).

The Avida System

The *Avida* system is a software platform wherein self-replicating computer programs ('digital organisms') adapt and evolve in a computational environment. Each digital organism has a genome comprised of a series of computer instructions, which, by default, are executed sequentially by a virtual CPU (central processing unit). However, some

instructions permit jumps or loops—for example, replication generally involves the execution of a copy loop. Execution of a viable genome results in an organism copying itself, instruction by instruction, and upon completion, dividing by binary fission to produce two organisms. If no empty space is available in the population, replication will result in the replacement of another organism in the population. Thus, the faster a given organism produces offspring, the more likely its genotype is to persist in the population over time.

Evolution occurs because the copy process is subject to random mutations, at a rate specified by the experimenter. Mutations can be point mutations, whereby one instruction in the genome is randomly replaced with another, or they can be insertions or deletions, which enable the genome to grow or shrink in length. Mutations are normally deleterious because they reduce the speed or efficiency with which an organism replicates; in the extreme, they are effectively lethal if they prevent an organism from being able to replicate altogether. Mutations that are beneficial increase the replication rate of the organism, either by improving the efficiency with which it produces copies of itself, or else by enabling it to receive additional CPU cycles, which allow it to execute more instructions. CPU cycles can thus be thought of as energy in *Avida*: every instruction executed burns a CPU cycle, but organisms must execute instructions in order to replicate or perform other functions. Digital organisms will thus generally adapt in one of two ways. First, they may evolve to reduce the number of CPU cycles they require to produce an offspring—that is, to reduce their generation time. Alternatively,

they may evolve to obtain more CPU cycles (more energy), which may permit them to produce more offspring but may also increase their generation time.

Digital organisms can obtain additional CPU cycles by performing bitwise logic functions on numbers they input from their environment. A correct computation will provide an organism with CPU cycles above its initial allotment, which can then be put toward further execution of the genome, potentially resulting in an increased rate of replication. Whether a given organism actually replicates faster depends on whether the CPU cycles obtained offset the additional CPU cycles required to perform the computation. Thus, organisms not only evolve to perform computations, but also to perform them as efficiently as possible.

The performance of computations represents a kind of a metabolism, in that the conversion of one or two numbers into another number provides the organism with energy. For the purposes of solving computations, organisms have a single genomic instruction, called `nand`; this instruction enables them to perform the NAND ('not and') logic function, provided that the instruction is properly coupled to input-output instructions to obtain the numbers and output the result. All other computations can be constructed by combining multiple `nand` statements with various other instructions. In this way, digital organisms also resemble real computers, in that all computations performed by computers can be built out of combinations of nands (also referred to as 'nand gates').

The EQU Function. For the current study, we employed nine possible logic functions, eight of which require two-inputs – that is, two binary numbers input from the environment. These nine bitwise logic functions are as follows: NOT, NAND, AND, OR-NOT, OR, AND-NOT, NOR, XOR, and EQU. Computation of these functions has been described elsewhere (Lenski et al. 2003a), but for purposes of illustration, we describe in greater detail one of these functions, 'Equals' (EQU), which is the focus of the current study. EQU is a computation where, for any two inputs, the correct output contains a '1' ('true') at every site where the bits are identical, and a '0' ('false') at every site where the bits are not identical. For example:

```
Input A:  0 1 0 1 1 0 1 1 1 0 0 1 . . .
Input B:  0 0 1 1 0 1 0 1 1 1 0 1 . . .
A EQU B:  1 0 0 1 0 0 0 1 1 0 1 1 . . .
```

Thus, in an environment where EQU is rewarded, an organism that inputs A and B and outputs the above number would receive additional CPU cycles. Because most computations require the coordination of multiple steps, digital organisms must store and manipulate intermediates or partial results. For example, the performance of EQU requires combining a minimum of 5 NANDs and at least 19 instructions in total (Lenski et al. 2003a). Finally, CPU cycle rewards are determined simply by comparing an organism's inputs to its output, such that selection is based on the phenotype, not the genotype.

In a recent paper, Lenski et al. (2003) examined the evolutionary origins of the EQU computation from an ancestral organism that could perform no functions. They found that the EQU function has the properties of a complex feature: its performance required the coordinated execution of numerous interacting parts. Moreover, its evolutionary emergence required that other, simpler functions also be rewarded; these simpler functions can then serve as building blocks for such a complex function. Here we expand upon this work by examining specialization of the EQU function. Starting from "generalist" organisms (those that could perform a variety of computations, in addition to EQU), we examine their evolution in a narrow environment, where only EQU generates extra CPU cycles. We examine how these digital organisms adapt to their novel environment, the extent to which they evolve to be highly specialized, and the evolutionary processes that govern their transition from generalists to specialists.

Methods

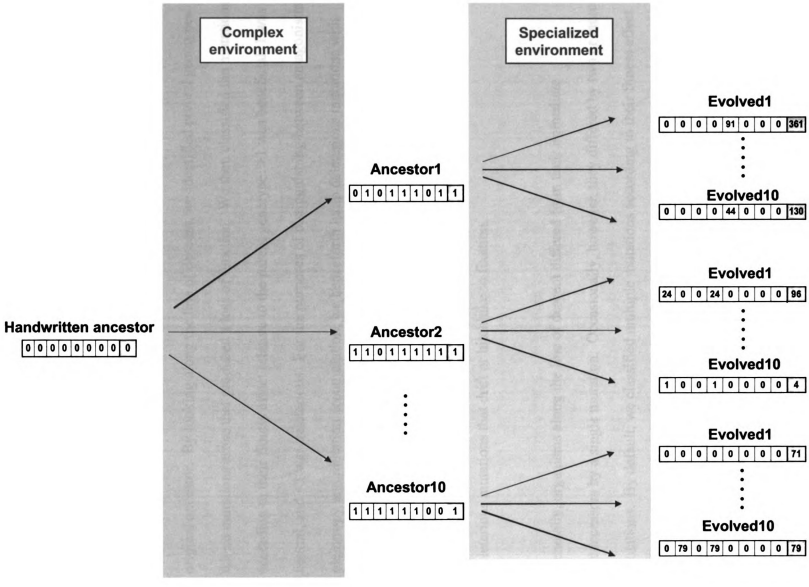
Experimental design

In the first stage of the experiment, replicate populations evolved from a single handwritten ancestor that could self-replicate, but could not perform any logic functions (Fig. 1). These populations evolved in an environment where the performance of all nine computations provided CPU cycles as rewards. These rewards were limited, however, to once per gestation cycle, such that organisms generally evolved to perform each function only once. (The gestation cycle is defined as the time from when an organism executes the first instruction in its genome to when it produces an offspring.) Following 100,000 updates of evolution, an arbitrary unit of time in *Avida* corresponding to an average of 30

instructions executed per organism (see Glossary; Appendix E), the dominant genotype was isolated from each population. These genotypes served as the generalist ancestors (subsequently denoted Ancestor1-Ancestor10) in the main experiment.

In the second stage of the experiment, replicate populations were founded from each generalist ancestor and evolved in an environment where only EQU yielded extra CPU cycles. The experiment consisted of 10 replicate populations for each of the 10 ancestors, for a total of 100 populations. The ancestors were generalists in that they could perform a wide variety of different logic functions, though they differed in the number and identity of the exact functions they performed (average = 7.3, range 6-9 of nine possible logic functions). All ancestors, however, performed EQU exactly once per gestation cycle. The ancestors also varied in the number of instructions comprising their genomes, with the shortest having 59 instructions and the longest having 124 instructions (average = 99.7). All populations evolved for 100,000 updates, during which time they only received a reward for the EQU computation. In this new EQU-only environment, however, organisms received rewards every time they performed the EQU computation and output the appropriate result. Insertion, deletion, and point mutations occurred at rates of 0.01, 0.01, and 0.08 mutations per genome per replication, respectively. Population size was limited to 3600 organisms, and the grid was started full, meaning that all positions were initialized with clones of the chosen ancestor. Offspring were placed randomly in the population, such that the population was well-mixed.

Figure 1. The evolution experiment had two phases: an initial period when replicate populations evolved in a complex environment with all functions rewarded, followed by a period of evolution in a specialized environment with only EQU rewarded. Evolved organisms represent the final dominant organism in each evolved population. Numbers shown beneath each organism specify its phenotype in terms of the number of times it performs each function during its life cycle, in the following order: NOT, NAND, AND, OR-NOT, OR, AND-NOT, NOR, XOR, and EQU. For clarity, the number that corresponds to an organism's performance of EQU has been shaded.


Complex environment
Specialized environment
Handwritten ancestor
0 0 0 0 0 0 0 0 0
Ancestor1
0 1 0 1 1 1 0 1 1
Ancestor2
1 1 0 1 1 1 1 1 1
Ancestor10
1 1 1 1 1 1 0 0 1
Evolved1
0 0 0 0 91 0 0 0 361
Evolved10
0 0 0 0 44 0 0 0 130
Evolved1
24 0 0 24 0 0 0 0 96
Evolved10
1 0 0 1 0 0 0 0 4
Evolved1
0 0 0 0 0 0 0 0 71
Evolved10
0 79 0 79 0 0 0 0 79

Examining the Line of Descent - To assess specialization and adaptive decay following evolution in the EQU-only environment, the most abundant genotype from each population was saved and assayed at the end of each experiment for its ability to perform each of the 9 computations, including EQU. For each of these genotypes, we also determined its line of descent, which is the sequence of all genotypes leading back to the original ancestor. By looking along the line of descent, we identified pivotal genotypes where mutations arose that produced a loss of function. We then classified the mutations according to their fitness effect relative to the parent genotype: >1 was beneficial, 1 was neutral, and <1 was deleterious. For the purposes of distinguishing between antagonistic pleiotropy and mutation accumulation, we henceforth lump deleterious mutations with neutral ones and refer to them collectively as non-beneficial. The reason for doing so is that the antagonistic pleiotropy hypothesis specifically concerns beneficial mutations, whereas mutation accumulation could encompass not only neutral mutations but also deleterious mutations that drift or hitchhike to fixation.

Generally, organisms along the line of descent differed from their immediate predecessors by a single mutation. Occasionally, however, they differed by two or more mutations. By default, we classified multiple mutations according to their fitness effect in combination. However, to ensure that these multiple mutational steps did not influence our results, we also analyzed our data without these multiple mutations. Finally, we repeated our experiments at higher and lower genomic mutation rates of 0.3 and 0.01, respectively, equal to 3-fold higher and 10-fold lower than our original experiments. To control for the effects of differential mutation supply, we performed

additional experiments where we scaled the length of the experiments inversely to the mutation rate. Thus, experiments at a genomic mutation rate of 0.3 were run for 33,000 updates, and those at the 0.01 genomic mutation rate were run for 1,000,000 updates.

Statistical Analyses

Performance of the EQU function - To determine how fitness and the performance of EQU varied depending on the ancestor, we performed two one-way ANOVAs. These analyses were performed using PROC GLM in SAS, with ancestor designated as a random effect. In the first ANOVA, we used the log relative fitness of evolved populations as the response variable, where each evolved population's fitness is relative to that of its own ancestor. In the second ANOVA, the response variable was the number of times EQU was performed in the numerically dominant genotype isolated from each evolved population. Because variances were heterogeneous across ancestors, we performed the ANOVAs as nonparametric Kruskal-Wallis tests. These tests were performed in SAS using PROC NPAR1WAY.

Antagonistic Pleiotropy versus Mutation Accumulation - To examine the relative contributions of antagonistic pleiotropy and mutation accumulation, we totaled the number of beneficial and non-beneficial mutations per ancestor across the ten replicate experiments at those steps where some unused function was lost. Because non-beneficial mutations are typically more common than beneficial mutations even in the line of descent (Lenski et al. 2003a), we also assembled a baseline calculation of the relative proportion of these two mutation types over the course of evolution by totaling their

number over the line of descent as a whole, irrespective of whether they were associated with a loss of function. To determine whether the ratio of beneficial to non-beneficial mutations was significantly higher among those mutations that caused a loss of function (which would provide support for the antagonistic pleiotropy hypothesis), we performed a Fisher's exact test that compared the number of beneficial versus non-beneficial mutations causing a loss of function to the number that did not. To assess the statistical significance of the contingency tables, we used the right-tailed p-value of a Fisher's exact test, where a low p-value would indicate that beneficial mutations were significantly overrepresented among mutations causing losses of function. The analyses were performed in SAS, using PROC FREQ and the Fisher Exact option.

Results

Specialization and Adaptive Decay in the EQU-only Environment

We consider three components of specialization. First, we examine the extent to which populations evolve increased performance of EQU, where the performance is determined as the total number of times an organism outputs the result of the EQU function per reproductive cycle. Second, because organisms can make improvements in the efficiency of their EQU performance, without increasing the number of times it is performed, we also consider the degree to which fitness increased in the EQU-only environment. Third, we examine the extent of adaptive decay—that is, the extent to which unrewarded functions were lost during evolution in the EQU-only environment, leading to the evolution of narrow niche breadth.

With regard to the first of these criteria, we find that evolved populations had greatly improved performances of EQU. Whereas all ancestors performed EQU only once per reproductive cycle, most evolved organisms performed it tens or even hundreds of times (Fig. 2). Interestingly, the magnitude of this improvement depended strongly on the ancestor (Kruskal-Wallis $\chi^2 = 42.41$, df = 9, $P < 0.0001$). There was also variation in the performance of EQU among replicate populations evolved from the same ancestor. For example, in 5 of 100 populations (three derived from Ancestor1 and one each from Ancestor9 and Ancestor10), the performance of EQU did not increase above the ancestral level. However, when averaged over the 10 replicate populations, organisms evolved from Ancestor1 had the third highest performance of EQU overall (Fig. 2). This result indicates that the chance occurrence of different mutations in replicate populations was an important component of specialization in this system. Similarly, because the generalist ancestors themselves evolved from the same handwritten ancestor (Fig. 1), differences in outcome that were contingent on the generalist ancestor also reflect the importance of chance events at an earlier stage of evolution.

While EQU performance did not increase in every population, fitness universally improved in the EQU-only environment (Fig. 3A). Once again, the magnitude of the increase depended greatly on the ancestor. Figure 3A shows the fitness trajectory of the populations over time, averaged over the 10 replicates for each ancestor. While all populations increased in fitness, there was substantial heterogeneity in the magnitude of

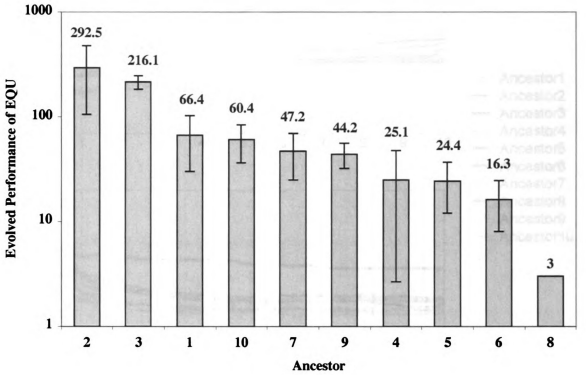


Figure 2. Average number of times EQU is performed per life cycle by evolved organisms, arranged from highest to lowest for each of the 10 ancestors. The ancestors all performed EQU only once, and each bar represents the mean across 10 replicate populations evolved from that ancestor. Values are plotted on a logarithmic scale, and error bars represent one standard error. For clarity, the average value is also shown above each bar.

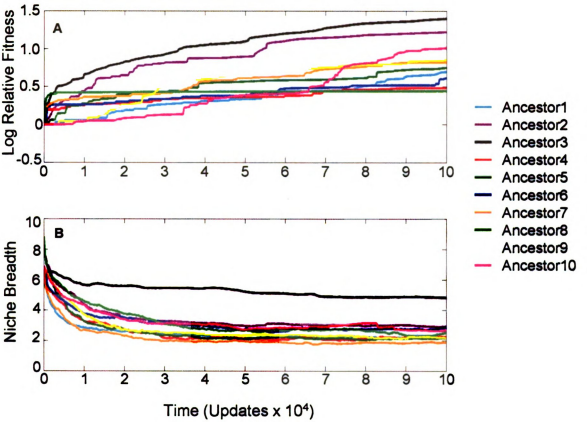


Figure 3. (A) Fitness trajectory of populations in the EQU-only environment. Each line represents the average of 10 replicate evolution experiments for each of 10 different generalist ancestors. Fitness is expressed as the log ratio of evolved relative to its own ancestor, such that all populations start at zero. (B) Reduction in niche breadth during evolution in the EQU-only environment. Niche breadth was calculated as the proportion of organisms performing each function at a given time and summed over all functions. Each line represents the average of 10 replicates for each of the 10 ancestors. Images in this thesis are presented in color.

this improvement, and the ancestor was a highly significant effect (Kruskal-Wallis χ^2 = 32.45, df = 9, P = 0.0002).

Evolution of Niche Breadth Reductions

The loss of unrewarded functions was not universal and also variable across ancestors. Failure to lose a particular function is indicated by a black cell in Figure 4. Only 7 of 100 populations retained only EQU and lost all unused functions; these 7 populations were distributed across 6 different ancestors (Fig. 4). Qualitatively, there was no association between losses of function and enhanced performance of EQU. For example, populations evolved from Ancestor3 tended to maintain a relatively broad niche (Fig. 4) and yet were the second-highest performers of EQU (Fig. 2). The decline in niche breadth over time is plotted in Figure 3B. Each line represents the average for a different ancestor, and the colors for each population are the same in the top and bottom graphs.

Population Genetic Processes Underlying the Evolution of Reduced Niche Breadth

Where functions were lost, we were interested in determining whether losses were caused by mutation accumulation or antagonistic pleiotropy. To address this question, we determined the fitness effect of every mutation that resulted in a loss of function along the line of descent. If mutations causing functions to be lost were neutral or deleterious in the EQU-only environment, it would indicate that mutation accumulation was

responsible for losses of function. Similarly, if the mutations leading to losses of function were beneficial, it would indicate that antagonistic pleiotropy was responsible. Note that there at least two ways in which a mutation causing a loss of function could be beneficial. First, it may be beneficial because the instructions encoding EQU also encode other functions, such that mutations that enhance EQU performance interfere with these other functions; this would constitute a classic example of pleiotropy. Second, mutations causing losses of function could also be beneficial because they reduce the energy spent performing useless tasks, thereby increasing fitness in the EQU-only environment. This also constitutes antagonistic pleiotropy, in the sense that a single mutation improves fitness in one environment, but reduces it in another. For our purposes, we did not distinguish between these two explanations: all mutations that were simultaneously beneficial in the EQU-only environment and resulted in a loss of function were interpreted as support for the pleiotropy hypothesis. Our classification scheme thus captures two categories of explanation: those mutations that fix via selection for improved performance in an EQU-only environment, and those mutations that fix by genetic drift or by hitchhiking alongside beneficial mutations.

The mutations leading to losses of function are shown for all populations in Figure 4, arranged by ancestor. Because a single mutation occasionally led to the simultaneous loss of multiple functions, cells are not necessarily independent of one another. In addition, because we are interested in understanding the niche breadth of the final derived organisms and the mutations that led to that state, we do not consider cases where a function was lost but subsequently regained. Thus, we only examine the mutations

causing losses of function if the function was absent at the end of the experiment. In cases where a function was lost, regained, and subsequently lost again, we consider only the final loss of function. This methodology is most likely conservative with respect to detecting antagonistic pleiotropy, as earlier mutations (when adaptation is most rapid) are more likely to be beneficial than later mutations.

For 8 of 10 ancestors, the beneficial to non-beneficial ratio was higher among mutations causing losses of function than among those that did not. In 7 of these 8 cases, the Fisher's exact test was highly significant (Table 1). In the two cases where the mutations causing losses of function were disproportionately neutral or deleterious (Ancestors 2 and 7), the differences were quite small. In these cases, the left-hand p-values of the Fisher's exact tests, which would test for overrepresentation of neutral or deleterious mutations among mutations causing losses of function, were nonsignificant ($P = 0.07$ and 0.53 for Ancestor2 and Ancestor7, respectively). Thus, these results show that where losses of function occurred, they were disproportionately likely to be caused by a beneficial mutation. This result implies that antagonistic pleiotropy was an important factor in driving the decay of unrewarded functions.

Steps with Multiple Mutations

Because some steps along the line of descent occasionally included multiple mutations (that is, a derived genotype differed from its immediate parent by more than a single mutation), we sought to determine whether these multiple mutations made a significant

Figure 4. Outcome of evolution in the specialized EQU-only environment for all 100 populations, arranged by ancestor. Each row specifies the final dominant genotype from one evolved population, and the colors indicate which functions were lost and the type of mutation (beneficial, neutral, or deleterious) that caused the loss of function. In some cases, a single mutation led to the loss of multiple functions at once, such that the colors of the blocks are not necessarily independent in every row. Images in this thesis are presented in color.

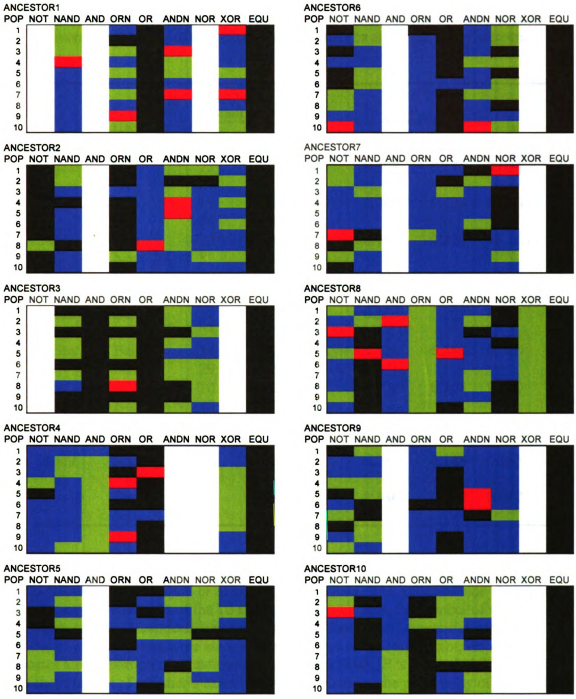

ANCESTOR1
ANCESTOR2
ANCESTOR3
ANCESTOR4
ANCESTOR5
ANCESTOR6
ANCESTOR7
ANCESTOR8
ANCESTOR9
ANCESTOR10
POP NOT NAND AND ORN OR ANDN NOR XOR EQU
1
2
3
4
5
6
7
8
9
10
KEY to CHART:
Organism did not lose this function
Ancestor did not perform this function
Mutation leading to loss of function was NEUTRAL
Mutation leading to loss of function was BENEFICIAL
Mutation leading to loss of function was DELETERIOUS

Table 1. Comparison of mutations associated with losses of function relative to their overall probability of occurrence, as substitutions along the line of descent.

Ancestor	Beneficial (Ben) or Not	Number of mutations: Loss	Not	P_B Loss	Not	*P*
Anc1	Ben	12	301	0.308	0.176	0.034
	Not	27	1407			
Anc2	Ben	15	592	0.306	0.424	0.966
	Not	34	803			
Anc3	Ben	20	496	0.714	0.370	<0.001
	Not	8	844			
Anc4	Ben	19	233	0.396	0.261	0.032
	Not	29	661			
Anc5	Ben	22	303	0.386	0.205	0.002
	Not	35	1178			
Anc6	Ben	14	268	0.326	0.183	0.020
	Not	29	1195			
Anc7	Ben	10	282	0.200	0.208	0.610
	Not	40	1076			
Anc8	Ben	28	201	0.431	0.172	<0.001
	Not	37	968			
Anc9	Ben	10	204	0.213	0.168	0.265
	Not	37	1010			
Anc10	Ben	17	394	0.386	0.232	0.017
	Not	27	1302			

Table 1. Results of Fisher's exact tests comparing the loss of functions due to beneficial versus non-beneficial (neutral or deleterious) mutations. The left-hand side shows the contingency table for each of the 10 ancestors. In each case, the number of mutations was summed over 10 replicate populations. "Loss" and "Not" categories refer to the number of mutations that were associated with a loss of function or not, respectively. P_B indicates the proportion of mutations that were beneficial. *P* is the probability associated with the right-tail of a Fisher's exact test—in other words, the probability of seeing, by chance alone, as much or more overrepresentation of beneficial mutations among loss of function mutations.

contribution to losses of function. We found that multiple mutations accounted for approximately 6.2% of all genotypic steps along the line of descent and for approximately 4.1% of mutations causing losses of function. Thus, multiple mutations were, if anything, underrepresented among the mutations causing losses of function. The Fisher's exact tests comparing losses of function due to beneficial versus nonbeneficial mutations (e.g., Table 1) were largely unaffected by the exclusion of multiple mutations (data not shown). The statistical significance of the results differed only for Ancestor1, which became nonsignificant once these mutational steps were excluded.

Niche Breadth Reductions at Higher and Lower Mutation Rates

Our initial experiments were performed at a genomic mutation rate of 0.1 for 100,000 updates. To assess the generality of these results, we repeated our experiments at significantly higher and lower mutation rates of 0.3 and 0.01, respectively. As expected, niche breadth usually declined more rapidly with increasing mutation rate (Fig. 5). However, it was not obvious whether the faster decay of niche breadth was a result of the greater overall mutation supply, or whether mutation rate disproportionately affected losses of function by altering the relative importance of beneficial and non-beneficial mutations. For example, in asexual organisms, increasing the mutation rate is expected to increase the fixation of non-beneficial mutations to a greater extent than beneficial mutations because, at high mutation rates, beneficial mutations will more often arise in different lineages that interfere with each other's fixation, a phenomenon termed

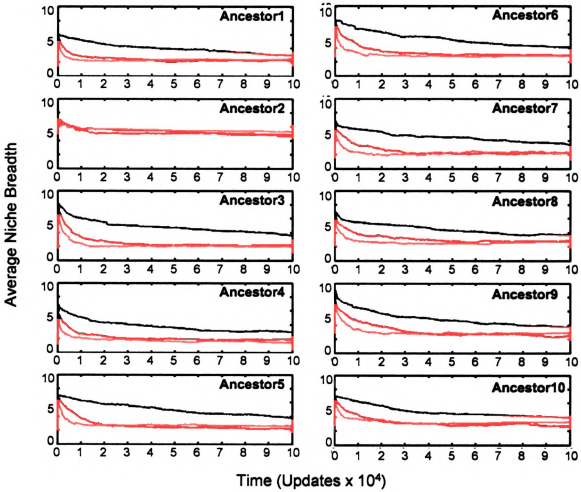


Figure 5. Decline in average niche breadth over time as a function of mutation rate. Average niche breadth was calculated as the mean niche breadth of the 10 replicate populations derived from each ancestor. Niche breadth was calculated as the proportion of organisms performing each function at a given time point and summed over all functions. Green: low, 0.01 genomic mutation rate; blue: medium, 0.1 genomic mutation rate; red: high, 0.3 genomic mutation rate. Images in this thesis are presented in color.

"clonal interference" (de Visser et al. 1999; Gerrish and Lenski 1998; Muller 1932; Orr 2000). The relative roles of antagonistic pleiotropy and mutation accumulation may thus be altered by changes to the mutation rate.

To address this issue, we repeated our experiments, but this time scaled their duration inversely to the mutation rate. Because our initial experiments were run at a 0.1 genomic mutation rate for 100,000 updates, we re-ran the high mutation rate (0.3) experiments for 33,000 updates, and the low mutation rate (0.01) experiments for 1,000,000 updates. Scaling the runs in this way is expected to control for differential mutation supply; this prediction was verified by examining the number of genotypes along the line of the descent, which was found to be similar across treatments (mean: low = 147.1, medium=141.9, high=139.4; all pairwise comparisons not statistically significant).

Our results show that increasing the mutation rate, while holding mutation supply constant, tends to decrease the number of beneficial mutations that cause losses of function per experiment (least squares means: low = 1.92, medium = 1.67, high = 1.38). A two-way parametric ANOVA based on the number of beneficial mutations causing losses of function found a significant effect of ancestor ($F_{9,270} = 12.41$, $P < 0.0001$), mutation rate ($F_{2,18} = 5.60$, $P = 0.013$), and their interaction ($F_{18,270} = 1.66$, $P = 0.046$). As expected, losses of function resulting from non-beneficial mutations showed the opposite pattern (least squares means: low = 2.76, medium = 3.03, high = 3.20). Ancestor and mutation rate were again both statistically significant ($F_{9,18} = 20.09$, $P <$

0.0001 and $F_{2,18} = 3.70$, $P = 0.045$, respectively). The interaction between mutation rate and ancestor, however, was nonsignificant ($F_{18,270} = 1.08$, $P = 0.376$) for these mutations.

Finally, we can ask whether beneficial mutations remain overrepresented among mutations causing losses of function at higher and lower mutation rates. These data are presented in Table 2, which shows the proportion of beneficial mutations, relative to the total, that were associated or not associated with a loss of function. At all three mutation rates, beneficial mutations were usually present in greater proportions among mutations causing losses of function than among those that did not. For all mutation rates, this proportion was higher in descendents of 8 out of 10 ancestors, although the identities of these eight ancestors varied across the treatments. Many of these differences were significant when we performed the Fisher's exact tests to examine the number of beneficial versus nonbeneficial mutations causing losses of function or not, as we also saw for our earlier analysis at the genomic mutation rate of 0.1 (Table 1). While the pattern at higher and lower mutation rates is qualitatively similar to that for the medium mutation rate, somewhat fewer tests were significant at both extremes. In general, however, the pattern was similar, despite large changes to the mutation rate, indicating that antagonistic pleiotropy was an important contributor to niche specialization at all three mutation rates.

Table 2: Proportion of mutations along the line of descent that were beneficial as a function of mutation rate.

	Genomic Mutation Rate:								
	Low (0.01)			Medium (0.1)			High (0.3)		
Ancestor	Loss	Not	*P*	Loss	Not	*P*	Loss	Not	*P*
1	0.162	0.195	NS	0.308	0.176	*	0.175	0.190	NS
2	0.321	0.333	NS	0.306	0.424	NS	0.288	0.252	NS
3	0.862	0.321	***	0.714	0.370	***	0.500	0.256	*
4	0.596	0.484	NS	0.396	0.261	*	0.426	0.199	***
5	0.545	0.267	***	0.386	0.205	**	0.354	0.363	NS
6	0.362	0.251	NS	0.326	0.183	*	0.341	0.200	*
7	0.192	0.186	NS	0.200	0.208	NS	0.192	0.175	NS
8	0.431	0.259	**	0.431	0.172	***	0.441	0.148	***
9	0.396	0.210	**	0.213	0.168	NS	0.227	0.226	NS
10	0.381	0.249	*	0.386	0.232	*	0.250	0.242	NS

Table 2. Comparison of the proportion of mutations substituted on the line of descent that were beneficial among those causing losses of function versus those that did not, at three different mutation rates. Notice that beneficial mutations are generally present in higher proportions among mutations causing losses of function. Asterisks indicate the significance of the associated Fisher's exact test, which compared the number of beneficial versus non-beneficial (neutral or deleterious) mutations that caused losses of function versus those that did not. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, NS not significant.

As before, we also analyzed our results to determine whether they were affected by the presence of multiple mutations in some steps. At the low mutation rate, multiple mutations comprised 2.3% and 0% of steps along the line of descent and losses of function, respectively. The conclusions of the Fisher's exact tests that compared the fitness effects of mutations causing losses of function to those that did not were generally unaffected by these mutations, with the exception of Ancestor4, for which the test became significant once steps with multiple mutations were excluded. At the high mutation rate, multiple mutations comprised 7.1% of all mutations along the line of descent, and 10.9% of mutations causing losses of function. The statistical significance

of all of the Fisher's exact tests at the high mutation rate were unchanged by the exclusion of these mutations.

Functional Genetic Explanations for Niche Conservatism

A striking feature of these experiments is the extent to which some functions were repeatedly retained across replicate populations started from the same ancestor (columns of black cells in Figure 4). For example, all ten populations evolved from Ancestor1 invariably retained OR, while those evolved from Ancestor3 always kept both AND and OR. There are at least two explanations for the maintenance of unrewarded functions. One possibility is that there may have been insufficient mutational pressure to cause losses of function. While this effect may be expected to be random with respect to functions, some functions present larger targets for mutations because they require more instructions to encode, and thus they may be lost more consistently. To test whether mutational pressure was strong enough to lead to decay of functions, we ran additional experiments with one ancestor, Ancestor3, for which derived populations had decayed the least on average over the course of their evolution. These experiments were identical to the original experiment, except that no functions—including even EQU—were rewarded. In ten replicate experiments starting from Ancestor3, every function was lost, showing that insufficient mutational pressure could not explain the failure for losses of function to occur.

A second possibility is that these functions were maintained because their performance was coupled to that of EQU—in other words, due to pleiotropy. One line of evidence that pleiotropy was often responsible for the maintenance of some functions is that their performance, despite not being rewarded, often increased during evolution in the EQU environment, and in many cases, in proportion to that of EQU. Figure 6 shows the phenotype of evolved organisms from three different ancestors. Correlations comparing the performance of retained functions to that of EQU are consistently significant (Ancestor10, OR-NOT: $r = 0.72$, d.f. = 6, $P = 0.044$; Ancestor1, OR: $r = 0.81$, d.f. = 8, $P = 0.005$; Ancestor2, NOT: $r = 0.98$, d.f. = 7, $P < 0.0001$; Ancestor2, OR-NOT: $r = 0.90$, d.f. = 6, $P = 0.002$).

Given that different functions appeared to be coupled in their performance, we wanted to see if we could understand the mapping between genotype and phenotype that gave rise to these correlations. In other words, rather than merely observing that some genotypes retained more unused functions than others, we wanted to understand the origins of this evolutionary pattern by investigating the relationship among different functions in the ancestral genome. To do so, we systematically assessed ancestral genomes for the extent to which knocking out a given instruction affected the performance of each function. The resulting "genotype-phenotype map" allowed us to infer the regions of the genome that encoded each function an organism performs, as well as the overlap in these regions, as described below.

Figure 6. Functions that were not lost were correlated with the performance of EQU. Each row represents the phenotype of a different evolved organism (one from each of the 10 replicate populations) for three illustrative ancestors. Numbers show the number of times the organism performs each logic function per life cycle. While many functions were lost (indicated by a zero), those that were not lost show a correlated increase in their performance with EQU. Correlation coefficients for the performance of these functions and EQU are indicated below each table. Correlations were calculated only between pairs of data where the function in question had been maintained (i.e., if the value in the table was greater than zero) in over half the replicate populations, as indicated by shading.

Ancestor 10

NOT	NAND	AND	OR ~	OR	AND ~	NOR	XOR	EQU
0	0	0	0	0	0	0	0	71
0	32	0	32	0	0	0	0	62
0	0	0	66	0	0	0	0	261
0	9	0	0	0	0	0	0	31
0	17	0	17	0	16	0	0	17
0	16	0	16	0	0	0	0	46
0	0	0	1	0	0	0	0	1
0	0	0	24	24	19	0	0	24
0	0	0	12	0	0	0	0	12
0	79	0	79	0	0	0	0	79

correlation with EQU: 0.72

Ancestor 1

NOT	NAND	AND	OR ~	OR	AND ~	NOR	XOR	EQU
0	0	0	0	91	0	0	0	361
0	0	0	1	1	0	0	0	1
1	0	0	0	1	0	0	0	3
0	0	0	0	101	0	0	0	101
0	0	0	0	32	0	2	0	62
0	0	0	0	1	0	0	0	2
0	0	0	0	1	0	0	0	2
0	0	0	0	1	0	0	0	1
0	0	0	0	1	0	0	0	1
0	0	0	0	44	0	0	0	130

correlation with EQU: 0.81

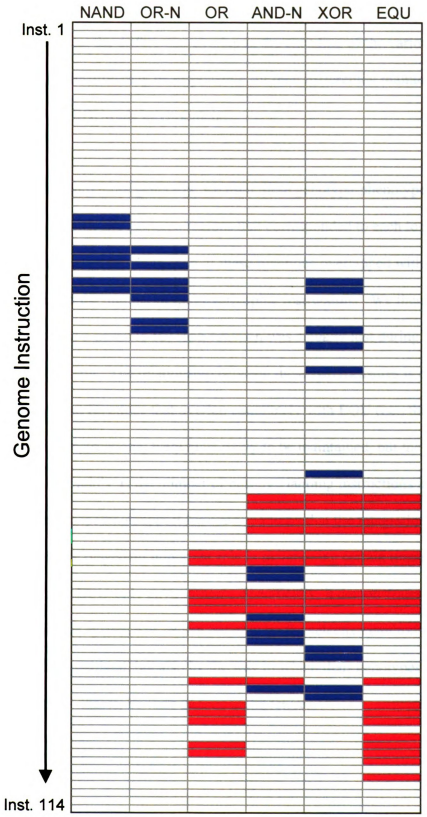
Ancestor 2

NOT	NAND	AND	OR ~	OR	AND ~	NOR	XOR	EQU
24	0	0	24	0	0	0	0	96
158	0	0	239	0	76	158	0	316
488	0	0	0	0	0	0	0	1952
1	48	0	49	0	0	0	0	196
1	0	0	1	0	0	0	0	4
11	0	0	11	0	0	0	0	64
58	0	0	58	0	0	0	0	173
0	1	0	1	0	0	0	0	4
29	0	0	0	0	0	0	0	116
1	0	0	1	0	0	0	0	4

correlation with EQU: 0.98 (NOT)
0.90 (OR~)

An example of a genotype-phenotype map, constructed for Ancestor1, is shown in Figure 7. Each row of the map represents one of the instructions in the genome, starting from the first (top row) to the last instruction (bottom row). Taking each site in the genome in turn, we replaced the instruction present at that site with a null instruction, called `nop-X`, and then tested the ability of the resulting "knock-out" mutant to perform logic functions. Organisms were only tested for functions that the unmutated "wild-type" organism had itself performed. Each column of the map denotes a different logic function that could be performed by the unmutated organism, and the cells are colored as follows. White means that when the instruction in the corresponding row is replaced with a null instruction, there is no effect on the function in the corresponding column. Colored cells indicate that replacing the corresponding instruction with a null instruction resulted in a loss of that function, and thus these cells correspond to the areas of the genome that encode the different functions. Among the colored cells, red cells represent the subset of the instructions required for any given function that are also required for the EQU function. For any other function, a mixture of red and blue therefore indicates only partial overlap with the instructions that encode EQU. Logic functions that lack any blue coloring, such as the function OR in Figure 7, indicate that there no sites in the genome that can be mutated to cause the loss of that function and still maintain the EQU function. We therefore expect that functions such as these might be rarely lost in an EQU-only environment, owing to a lack of a mutational target that does not also affect EQU. Consistent with this expectation, populations evolved from this ancestor in an EQU-only environment never lost the ability to perform OR, but usually lost all other unnecessary functions (Fig. 4, upper left panel). Analysis of the genotype-phenotype map for this

Figure 7. Genotype-phenotype map showing the instructions that encode each function in Ancestor1. Each row in the map represents a single instruction, starting from the first instruction in the organism's genome (top row) to the final instruction (bottom row). Each column represents a different function performed by the ancestor, and the coloring indicates what happens to the performance of that function when a given instruction is knocked out (replaced by a null instruction). White: knocking out the instruction does not affect performance of the function. Blue or Red: knocking out the instruction causes the function to be lost. Red blocks indicate the subset of instructions required for a given function that, when knocked out, also cause the loss of EQU. Note that every instruction in this organism that knocks out OR also knocks out EQU, whereas this pattern does not hold for NAND, OR-N, AND-N, or XOR. Images in this thesis are presented in color.


NAND
OR-N
OR
AND-N
XOR
EQU
Inst. 1
Genome Instruction
Inst. 114

ancestor thus implies that the differential overlap in the encoding of the various functions with that of the EQU function may be responsible for their differential maintenance during evolution in an EQU-only environment.

To assess this relationship more generally, we used these genotype-phenotype maps for all the ancestors to identify the genomic regions that corresponded to each logic function. We then assessed the extent to which each of these functions overlapped with the EQU function and calculated the number of non-overlapping instructions. We then determined how many times (out of a possible 10) each function was actually lost during evolution in an EQU-only environment. The relationship between these two measures is presented in Figure 8, and shows that functions that overlap completely with EQU (i.e., those with zero non-overlapping instructions) were most likely to be maintained, but the probability of maintenance drops rapidly as the number of non-overlapping instructions increases. This result provides compelling support for our hypothesis that the integration of these functions in the genome played an important role in maintaining certain unused functions during evolution in the EQU-only environment. Specifically, it demonstrates that niche breadth evolution in this system was driven not only by the selective environment in which these organisms evolved, but also by the way in which their genotypes mapped onto their phenotypes—that is, by their genetic architecture.

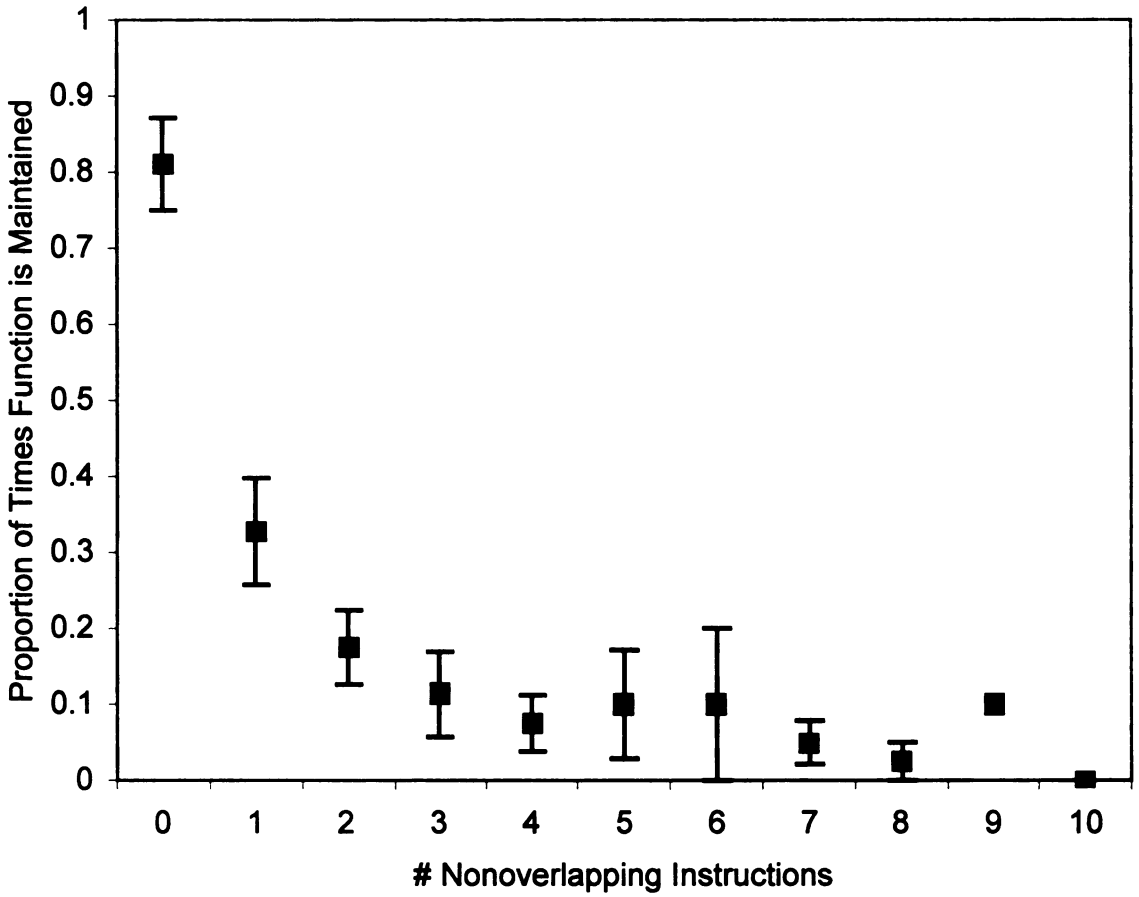


Figure 8. Association between the number of nonoverlapping instructions (required to perform some function but not required for the EQU function) and the proportion of times (out of a possible 10) that the function was maintained during evolution in the EQU-only environment. Error bars represent one standard error.

Discussion

Two distinct population genetic mechanisms are thought to promote the evolution of ecological specialization, reflected in a narrow niche breadth. One entails the accumulation of mutations that are neutral or deleterious in a novel environment owing to relaxed selection on unused functions. In other cases, fitness improvements in a novel environment may come at the expense of other traits, leading to trade-offs and losses of function. Trade-offs can occur if the same genes contribute to two or more traits, such that mutations that improve one may worsen others. However, even when traits do not share a genetic basis, trade-offs can still arise if the maintenance of unselected or weakly selected traits entails an energetic burden.

Here, we describe the evolution of ecological specialization in digital organisms. Starting from a set of generalist ancestors, each of which could perform a wide variety of logic computations, we examined their adaptation to a novel environment where only a single computation was directly selected. A benefit of examining the process of specialization in digital organisms is that we can precisely trace the mutational steps leading from the generalist ancestor to the evolved specialist, which allows us to examine in detail the mutations that lead to losses of function along the way.

Our results revealed significant heterogeneity in the magnitudes of fitness improvements in the novel EQU-only environment, with different populations evolving to perform EQU to different extents depending on the ancestor used to initiate the experiment. All of the ancestors performed EQU once per gestation cycle at the start of the experiment and, in a

few cases (5/100), evolved organisms did not increase their performance of EQU above the ancestral level. In most cases, however, organisms evolved to perform the function tens or even hundreds of times per gestation cycle. The evolved organisms also varied in the extent to which their niche breadth became narrower, with very few populations (only 7/100) evolving to become pure EQU specialists.

Examination of the mutations that led to losses of function allowed us to quantify the relative importance of mutation accumulation and antagonistic pleiotropy. These data showed that, in absolute terms, more losses of function were caused by neutral or deleterious mutations than by beneficial mutations. Yet, when we standardized for the greater numbers of non-beneficial substitutions along the lines of descent, beneficial mutations were disproportionately associated with losses of function. Although the proportion of losses of function that could be attributed to beneficial mutations was generally higher at lower mutation rates, the proportion of beneficial substitutions overall was also higher, such that changes to the mutation rate had little effect on the general results.

The finding that lower mutation rates permit the fixation of proportionally more beneficial mutations suggests that some kind of interference is occurring at the higher mutation rates, although it is not clear whether the interference arises from deleterious or beneficial mutations. At high mutation rates, beneficial and deleterious mutations may often arise on the same background, limiting fixation to those beneficial mutations of large effect (Johnson and Barton 2002; Orr 2000; Peck 1994). High mutation rates can

also lead to interference among beneficial mutations that arise in different clonal lineages (de Visser et al. 1999; Gerrish and Lenski 1998). Distinguishing between these two alternatives in evolving digital populations is a subject for further study.

One surprising result of these experiments was how, in particular ancestors, certain functions were often maintained in the absence of direct selection for their performance. Examination of the genetic architecture of the ancestors revealed that overlap in the genetic instructions that encode the different functions was a good predictor of their maintenance during evolution in the EQU-only environment. Not only were these functions maintained, but their performance also often increased in parallel with that of EQU, resulting in unexpected positive correlations between certain traits across populations evolved from the same ancestor. Because we know that there was no direct selection on these functions, their maintenance is more analogous to that of a vestigial trait, rather than the outcome of selection operating on two traits simultaneously. Wright (1977, p. 428) and Lande (1978) both suggested that useless or even slightly detrimental functions might be retained over long periods of time owing to their pleiotropic relationships to characters under direct selection. Nevertheless, it would be interesting to examine the consequences of these genetically integrated traits in the event that selection were to operate on them in opposing directions, and we will examine this possibility in future work.

Our results show that there was no single function that was always retained with EQU; rather the identity of the retained functions varied depending on the particular ancestor.

For example, organisms evolved from Ancestor3 failed to lose AND and OR, whereas organisms evolved from Ancestor2 often failed to lose NOT and OR-NOT. Moreover, because the ancestors all shared the same historical environment, these differences in outcome reflect stochasticity in the origins of each ancestor's unique genetic architecture—a genetic architecture that influenced the subsequent trajectory of evolution in the EQU-only environment. Where multiple functions were maintained (e.g., Ancestor3), it would be interesting to explore whether they had been built upon each other sequentially. One could imagine, for instance, that EQU evolved from AND, and that AND evolved from OR, and so on. Of course, the construction of complex functions out of simpler ones—a process that contributes to the emergence of pleiotropy in this system—also occurs in natural systems (Chen et al. 1997; Dean and Golding 1997; Jacob 1977; Meléndez-Hevia et al. 1996; Nilsson and Pelger 1994). Thus, investigations into the form and direction of pleiotropy in nature might be informed through a consideration of the evolutionary history of the traits in question.

The importance of genetic integration for the maintenance of unrewarded functions in highly specialized environments led to substantial variation in the niche breadth of evolved organisms, with some organisms evolving very narrow specialization, and others maintaining their niche breadth at about half their ancestral levels. This result raises interesting questions about the relative long-term success of these organisms in a fluctuating environment, where antagonistic pleiotropy and mutation accumulation may continually degrade functions that might become necessary again at some later time (Kawecki 2000). Organisms with highly integrated genetic architectures would

potentially prosper in such environments, whereas those with greater modularity might do better in a more stable environment, particularly if genetic correlations among traits were found to constrain the optimization of each trait individually. These predictions do not differ from existing theories about the kinds of environments that select for generalist versus specialist species, with the former predicted to emerge in a temporally heterogeneous environment, and the latter when there is environmental constancy (Futuyma and Moreno 1988; Kassen 2002; Levins 1968). However, this perspective emphasizes the role of genetic architecture in mediating these transitions, rather than selection as the sole determinant of niche breadth.

A common finding in these experiments, as well as in others that have employed digital organisms, is the prevalence of deleterious mutations along the line of descent, indicating that it is not uncommon for them to attain fixation in these populations. Several authors have recently considered the role of deleterious mutations in adaptation, and this work has led to a re-evaluation of how mutation rate alters the rate of adaptation in asexual organisms (Johnson and Barton 2002; Orr 2000; Wilke 2004). One of the difficulties encountered by this work is the complexity of the process, which requires modeling many competing lineages and evaluating non-equilibrium conditions, making it difficult to derive exact solutions. Digital systems may prove to be a suitable testing grounds for some of the hypotheses generated by this work, particularly because of the ease of estimating parameters that are difficult to measure in biological systems, such as the rate of occurrence and fixation of beneficial and deleterious mutations (see also Rozen et al. 2002; Sanjuan et al. 2004). Moreover, mutation accumulation explanations for

specialization often assume that the relevant mutations are either conditionally or weakly deleterious, because unconditionally deleterious mutations have difficulty attaining fixation except in small populations (Kawecki 1994; MacLean et al. 2004). In asexual organisms, however, deleterious mutations can hitchhike to fixation alongside beneficial mutations. Given that many extreme examples of adaptive decay involve bacteria (Cole et al. 2001; Ochman and Moran 2001; Wernegreen et al. 2002), the potential role of deleterious mutations needs to be considered more carefully. In asexual organisms, niche-breadth reductions could be occurring by both antagonistic pleiotropy (fixation of beneficial mutations) and mutation accumulation (via increased fixation of deleterious mutations), with the fixation of the former predisposing that of the latter through hitchhiking.

Ecological theories of niche specialization predict that organisms will evolve to match the heterogeneity of their environment (Kassen 2002; Levins 1968; Scheiner 1993; Via and Lande 1985). Our results show that environmental constancy can, in fact, drive the evolution of niche breadth reductions, with all organisms evolving niche breadths that were narrower than that of their ancestors. However, substantial diversity in niche breadth was observed among independently evolved organisms despite their evolution in identical environments. While the extent to which traits are encoded by the same or different genes has sometimes been taken into consideration when predicting the relative importance of antagonistic pleiotropy and mutation accumulation in driving specialization (Fry 1993; Futuyma and Moreno 1988; Kawecki 1994; Kawecki 1998), the degree to which suites of traits may be genetically integrated has not often been

considered with regard to the maintenance of functions across environments (but see Rausher 1988).

Although trade-offs are widely thought to promote the evolution of ecological specialization, the requisite negative correlations have often not been forthcoming (Agrawal 2000; Fry 1996; Jaenike 1990). The failure to detect negative correlations has led to a developing body of work that focuses on alternative explanations for the evolution of ecological specialization (Kawecki 1994; Kawecki 1998; Whitlock 1996). Also, Rausher (1988) suggested that trade-offs may not always be detected. For example, studies of diet breadth in phytophagous insects often employ host species that are already part of the natural diet. If severe trade-offs exist, then those host species are more likely to have been excluded from the diet previously, and the observed niche breadth will consist only of hosts for which there was either little or no conflict. Our results are consistent with this hypothesis, with trade-offs often quickly leading to losses of function, leaving mostly positive correlations among the remaining functions.

Ultimately, of course, experiments with digital organisms cannot tell us what processes are actually at work in any given natural system – that is an empirical question that cannot be addressed by any model system, digital or otherwise. However, digital systems provide a novel way of assessing the logic that underlies many evolutionary theories, especially where complex interactions limit the opportunity for purely theoretical analysis. Our results show that ecological specialization occurs in digital organisms and, moreover, that some of the same patterns that have complicated simple theories of niche

breadth in natural systems, such as the apparent paucity of trade-offs and an excess of positive correlations, also emerge here. Finally, digital systems offer the ability to connect patterns to processes, and thus allow investigations of causal mechanisms more directly than is possible in any other system, enabling tests of existing hypotheses as well as the development of new ones that can in turn be tested in other systems.

CHAPTER 4

CORRELATED TRAITS AND RUGGED ADAPTIVE LANDSCAPES IN DIGITAL ORGANISMS

Pleiotropy and epistasis lie at the heart of much of evolutionary theory. Both give rise to the complex mapping between genotype and phenotype and are hypothesized to generate constraints on adaptation. Pleiotropy is a major source of genetic correlations (with the other being linkage), which can hinder the response to selection on one trait owing to its correlation with another (Lande 1979; Via and Lande 1985). If sufficiently strong, pleiotropy can prevent traits from being independently optimized. However, even where pleiotropy does not confer an absolute constraint, it can still prolong an approach to the optimum and lead to temporarily maladaptive states.

Epistasis can also constrain evolution; epistasis for fitness in particular generates rugged adaptive landscapes, with their potential to trap populations on suboptimal fitness peaks (Whitlock 1996; Wright 1968). Despite the conceptual appeal of envisioning adaptation as a process that unfolds on rugged adaptive landscapes, their importance has been difficult to demonstrate (Coyne et al. 1997; Coyne et al. 2000, although see Korona et al. 1994). Even where there is evidence that populations reside on alternative peaks, it can be difficult to determine even the simplest attributes of the adaptive landscape. For instance, hybrids of low fitness may indicate an intervening valley, but whether the actual evolutionary trajectory involved traversing it is unknown. An alternative possibility is that the populations diverged around a ridge, such that the true adaptive landscape is volcano-shaped rather than comprised of two peaks (Dobzhansky 1937; Gavrilets and Hastings 1996). Demonstrating the importance of rugged adaptive landscapes for evolution clearly requires detailed knowledge of both the trajectory of the evolving populations and the fitness effects of the contributing mutations, but this is not feasible in most systems.

Digital evolution systems offer a unique opportunity to examine these processes. Several features of digital systems make them particularly well suited for addressing questions in the realm of evolutionary genetics. First, they share with other experimental evolution systems the benefits of large population sizes and short generation times, which permit substantial adaptation to occur over short time scales. Second, clones or even entire populations can be saved and restored at a later time point, permitting direct comparisons of evolved and ancestral genotypes. Third, it is possible to track the precise trajectory of

evolving populations, as well as to determine the fitness effects of all mutations that arise along the line of descent leading from the ancestor to the evolved organism. A complete battery of genetic tools and tests are available, including genome sequences, perfect phylogenetic reconstructions, and a map of the mutational neighborhood of any genotype of interest. Fourth, with 26 different instructions possible at any site, and viable genome sizes ranging from 12 to over 1,000 instructions in length, the number of possible genotypes in the system is vast. Evolution thus proceeds in genetically diverse populations that are potentially far removed from a state of equilibrium. Finally, perhaps the most important attribute of these systems is that of a complex, nonlinear mapping between genotype and phenotype, a property that permits the emergence of pleiotropy and epistasis. These final two qualities mean that digital organisms, in principle, may face many of the same complexities of adaptation that beleaguer their biological counterparts. The goal of the current study is to address this possibility and then to capitalize on some of the unique experimental capabilities of digital systems described above to shed light on these processes.

Below, we describe a set of experiments to examine the evolutionary response of genetically coupled traits in digital organisms. When multiple traits share an underlying genetic basis, they may not be able to evolve independently, resulting in constraints on adaptation, or even maladaptation (Conner 2003; Crespi 2000; Via and Lande 1985). Selection experiments can be useful for examining constraints that arise from pleiotropy (Barton and Partridge 2000; Beldade et al. 2002; Weber 1996). However, most studies rely on the detection of genetic correlations among traits, which are then used to draw

inferences about both the underlying genetic architecture of the traits, as well as their expected response to selection. Little is known about whether genetically coupled traits can be uncoupled by selection, nor whether they play a role in directing long-term macroevolutionary outcomes. For example, theory suggests that genetic correlations are unlikely to constrain evolution permanently in an environment with only a single optimum (Via and Lande 1985). Nevertheless, genetic correlations can be important in directing evolution toward a particular adaptive peak in a multi-peaked environment, and thus may lead to divergence over longer evolutionary times scales (Price et al. 1993; Schluter 2000; Steppan et al. 2002)

In digital organisms, the relevant "traits" are logic computations, which organisms evolve to perform using numbers they input from their environment. Computation of logic functions provides the organisms with additional energy, which they can use to reproduce. The specific computations we examine were identified in a previous study in which replicate populations evolved in an environment where only a single computation, the logic function EQU, was directly selected (Ch. 3). A major result of that work was that selection to increase outputs of the EQU computation often led to a correlated increase in the outputs of other, unselected computations. Moreover, the reason for the correlated increases became apparent when we examined the way in which different computations were encoded in the genome—functions prone to correlated increases also exhibited high degrees of genetic overlap with the selected EQU function (Figures 1 and 2). This result indicated that pleiotropy—in the sense that the same portions of the genome affected the expression of multiple logic functions—was the cause of their

Figure 1. Genotype-phenotype map for Ancestor1. Each row represents one instruction, starting from the first instruction in the organism's genome (top row) to the final instruction in the genome (bottom row). Columns indicate each logic function performed by the ancestor, and shaded cells indicate instructions that, when knocked out, result in a loss of the corresponding logic function. Knock-outs were performed by replacing the instruction present at the site with a placeholder instruction called `nop-X`, which has no function. Note that every instruction that necessary for EQU is also necessary for OR, whereas the reciprocal is not true.

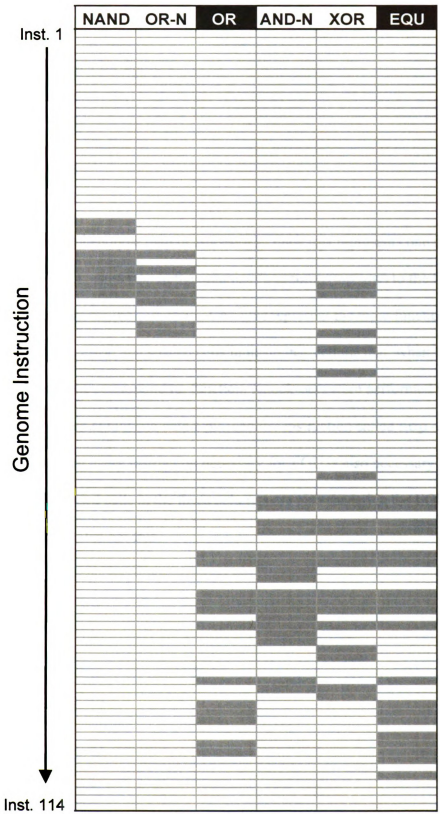
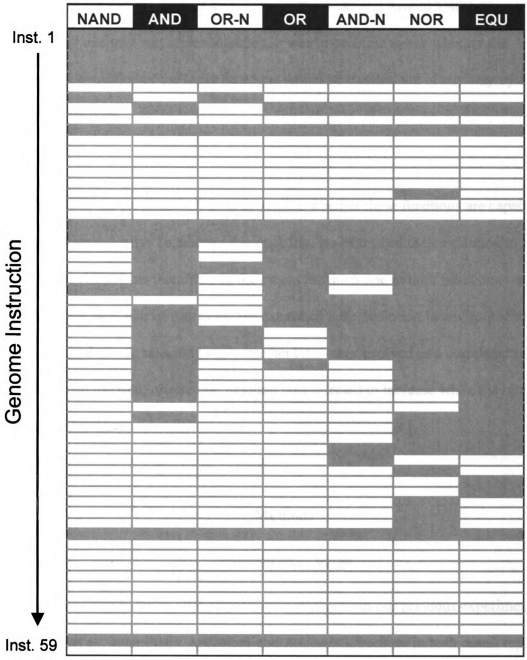

NAND
OR-N
OR
AND-N
XOR
EQU
Inst. 1
Genome Instruction
Inst. 114

Figure 2. Genotype-phenotype map of Ancestor3. Each row represents one instruction, starting from the first instruction in the organism's genome (top row) to the final instruction in the genome (bottom row). Columns represent the logic functions performed by the ancestor, and shaded cells indicate instructions that, when knocked out, result in the loss of the corresponding function. Knock-outs were performed by replacing the instruction present at that site with a placeholder instruction called `nop-X`, which has no function. Note that every instruction necessary for EQU is also necessary for OR and AND (that is, all shaded cells in the these columns are also shaded in the EQU column), whereas the reciprocal is not true.


NAND
AND
OR-N
OR
AND-N
NOR
EQU
Inst. 1
Genome Instruction
Inst. 59

coordinated evolution, a result consistent with the predications of quantitative genetics theory (Falconer and Mackay 1996; Lande 1979; Lande and Arnold 1983). However, because only one trait was under selection, it was impossible to say whether the independent evolution of the logic functions would be constrained in any way by their genetic association.

The purpose of the present study is to determine whether these functions are capable of evolving independently. To address this question, we examined their evolution in a variety of environments that differ in the extent to which one or both functions was under selection. For each pair of functions, we examined their evolution in environments where only one function was selected to increase, and the other evolved as a correlated response to selection on the first, or else one function was selected to increase while the other was selected to decrease.

Methods

Experimental Design

Ancestors and Traits Pairs—We chose two ancestors from our previous experiments (Ostrowski et al., submitted), Ancestor1 and Ancestor3, because in both cases, replicate populations evolved from these ancestors in an EQU-only environment (an environment where only computation of EQU yielded additional CPU cycles) always maintained the function OR, despite losing most other unrewarded functions. Not only did organisms evolved from these ancestors fail to lose OR, but its output often increased in a correlated

fashion with that of EQU, as measured across replicate populations (Table 1). Organisms evolved from Ancestor3 in an EQU-only environment not only showed correlated increases in OR, but also in the computation of AND (Table 1). For this reason, we also examined the relationship between these two functions in populations evolved from Ancestor3.

Table 1. Correlations among traits (number of times functions were performed during an organism's lifetime) observed in 10 replicate populations evolved from two ancestors, where selection was for the EQU trait only.

Ancestor1

Pop.	NOT	NAND	AND	ORN	OR	ANDN	NOR	XOR	EQU
1	0	0	0	0	91	0	0	0	361
2	0	0	0	1	1	0	0	0	1
3	1	0	0	0	1	0	0	0	3
4	0	0	0	0	101	0	0	0	101
5	0	0	0	0	32	0	2	0	62
6	0	0	0	0	1	0	0	0	2
7	0	0	0	0	1	0	0	0	2
8	0	0	0	0	1	0	0	0	1
9	0	0	0	0	1	0	0	0	1
10	0	0	0	0	44	0	0	0	130

correlation between OR and EQU: 0.806

Ancestor3

Pop.	NOT	NAND	AND	ORN	OR	ANDN	NOR	XOR	EQU
1	1	117	118	117	118	0	0	0	118
2	0	0	100	0	100	0	0	0	298
3	1	140	141	140	141	141	0	0	141
4	1	0	77	0	77	0	0	0	152
5	104	0	104	0	104	0	0	0	104
6	1	164	165	164	165	0	0	0	322
7	209	0	208	0	208	0	0	0	208
8	1	0	185	0	185	185	0	0	369
9	123	1	124	1	124	124	0	0	124
10	0	114	109	0	109	0	0	0	325

correlation between AND and EQU: 0.368
correlation between OR and EQU: 0.368

Evolution environments—We evolved populations founded with either Ancestor1 or Ancestor3 in each of four environments. In two of these environments, we rewarded the performance of one or the other function, while neither punishing nor rewarding the other. "Rewarded" functions are those that provide additional CPU cycles every time an organism outputs it, whereas "punished" functions are those computations that, when output, cause CPU cycles to be lost. Every time an organism outputs a correct computation, it receives additional CPU cycles, and each output of a computation is considered to be one "performance" of the computation. Because these additional CPU cycles are awarded without regard to how the organism performs the computation, selection is a function of the organism's phenotype, and not its genotype. The magnitudes of the various punishments and rewards used in this study are outlined in Appendix F.

For the purposes of assigning a name to each of these environments, we have adopted the convention of using a "+" in front of a function to indicate that it was rewarded in a particular environment and an "-" to indicate that it was punished. For example, in the case of EQU and OR, we evolved replicate populations in the following four environments:

(1) +EQU (environment that rewards EQU)

(2) +EQU/-OR (environment that rewards EQU but punishes OR)

(3) +OR (environment that rewards OR)

(4) +OR/-EQU (environment that rewards OR, but punishes EQU)

The experiments had 100-fold replication in each evolution environment, for a total of 1200 runs (3 ancestor-function pairs x 4 environments x 100 replicates). All experiments were run for a period of 100,000 updates at a genomic mutation rate of 0.1 divided among point, insertion, and deletion mutations, which occurred at rates of 0.08, 0.01, and 0.01 mutations per genome per generation, respectively. Placement of offspring was mass action, such that the populations were genetically well mixed. At the end of each experiment, the final dominant genotype was isolated from each population and assayed for its ability to perform the logic functions of interest. For each evolved organism, we also determined its line of descent—that is, the sequence of all genotypes, leading from the ancestor to the evolved organism. Examining the line of descent allows us to determine the evolutionary trajectory of a lineage over time, and therefore to identify where mutations arose that led to the loss of a function. Not only can we identify each mutation along the line of descent, but we can also determine its fitness effect in the environment in which it arose, as well as in any other environment that might be of interest. In the current study, we use this information to examine whether a given evolutionary trajectory, as it unfolded in one environment, would have been likely in another environment, and thus how changes in the topology of the adaptive landscape altered the outcome of evolution.

Results

Direct and correlated responses to selection on functions OR and EQU in Ancestor1

The results of experiments to examine the association between OR and EQU in Ancestor1 are shown in Figure 3. At the start of the experiment, the ancestor could

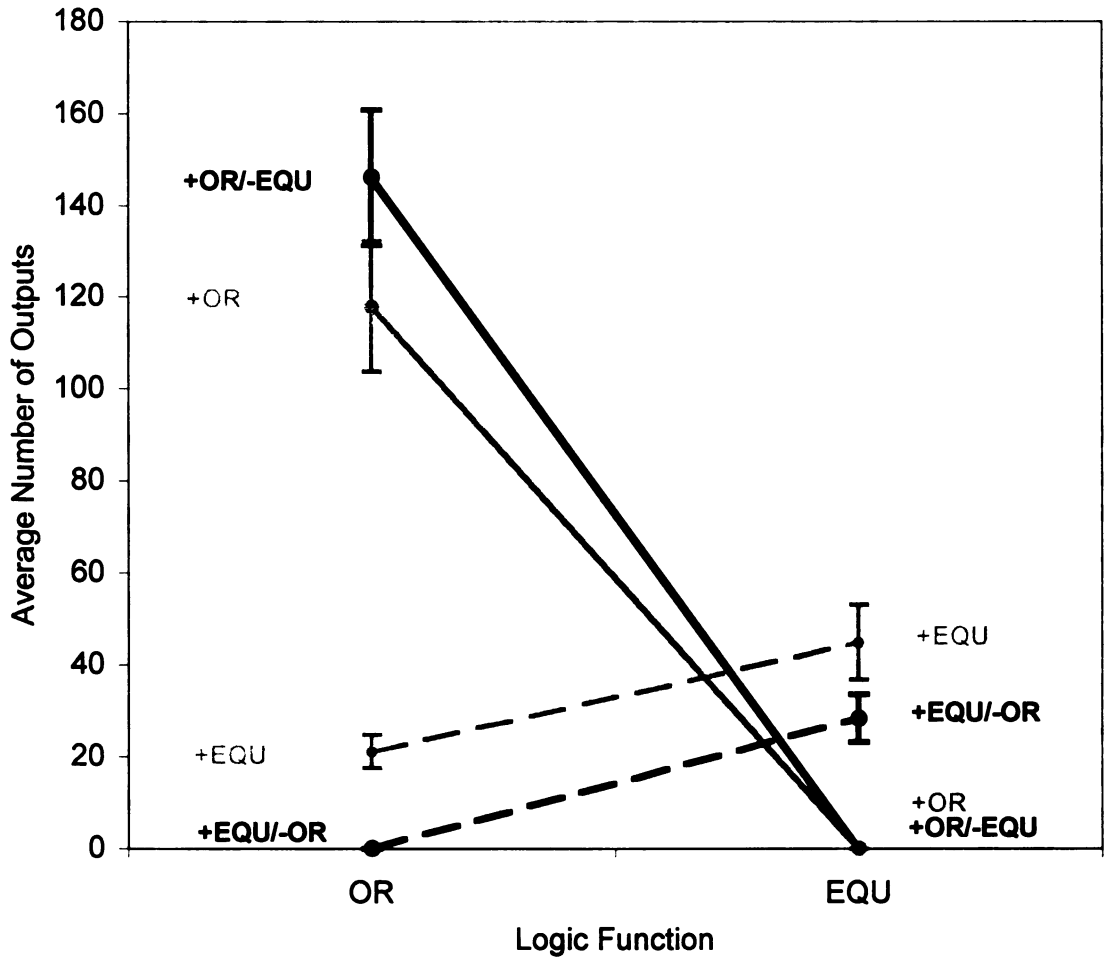


Figure 3. Number of times per reproductive cycle a given logic function is output, as a function of the environment in which the organisms evolve. Each point represents the mean output of a given function, determined as the performance of that function by the most common genotype in the population isolated at the end of the experiment. The evolution environment is indicated next to each point, and lines connect measurements made on the same set of populations. Error bars indicate one standard error, based on 100 replicates in each experiment.

perform each of these functions only once. Each point on the graph indicates the average number of times a given function was performed by 100 independently evolved organisms, depending on the environment in which they evolved. Starting on the left side of the graph, the results show that the evolved performance of OR is higher in the two environments where it was rewarded, +OR and +OR/-EQU, than in the two environments where it was either not rewarded or punished (Fig. 3; comparing upper left to bottom left). This result indicates that OR responds more strongly to direct selection. Of the two environments in which it was rewarded, the performance of OR was higher when EQU was punished than when it was not (mean = 146.07 in +OR/-EQU environment, versus mean = 117.90 in +EQU environment). However, this difference was not quite statistically significant (Mann-Whitney U = 5763, n = 200, $P = 0.062$). The lower lefthand portion of the figure indicates that the performance of OR also increased above the ancestral level of 1 as a correlated response to selection on EQU (mean OR = 21.1 in +EQU environment). However, when OR was punished, it was lost completely (mean OR = 0 in +EQU/-OR environment; Fig 3, lower left). This result indicates that, despite the correlated response of OR to selection on EQU, the association between the functions could be broken when selection acted on them in opposing directions.

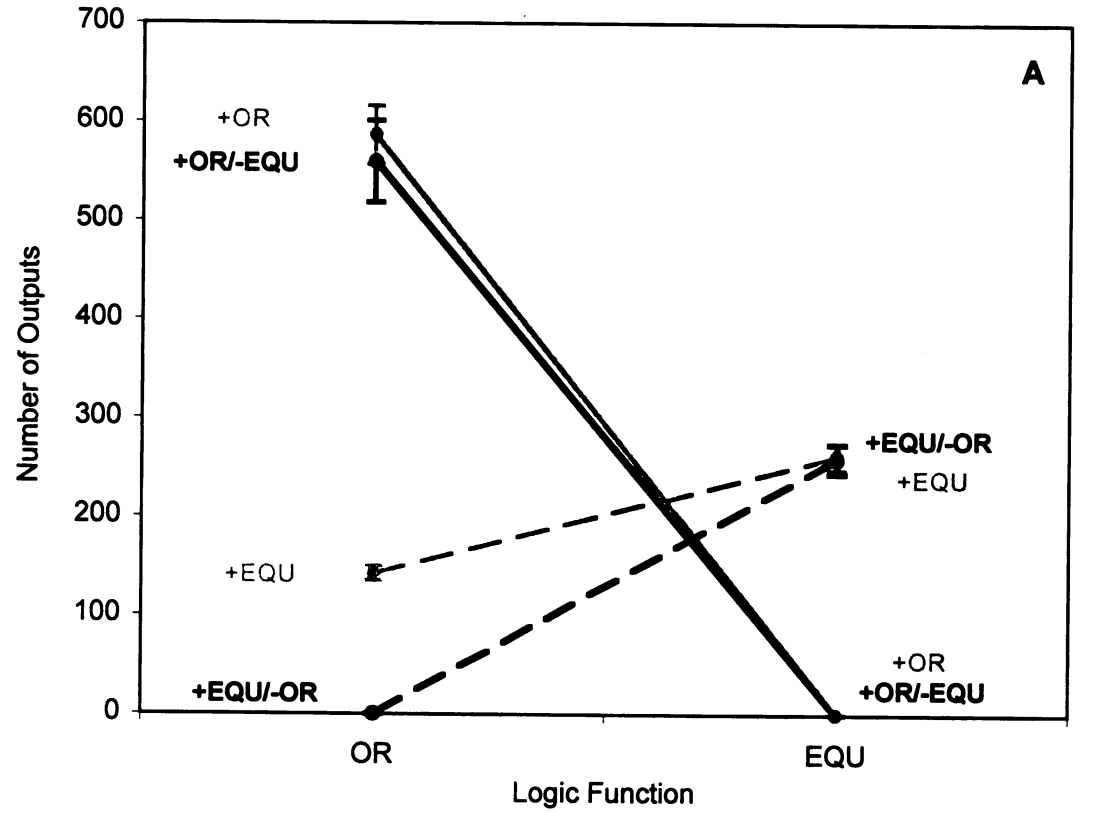
The response of EQU was in most respects similar to that seen with OR. First, its performance evolved to higher levels in the environments in which it was directly selected (i.e., in the +EQU and +EQU/-OR environments; Fig 3, lower right). The evolved performance of EQU also did not differ significantly depending on whether OR was being punished or not (Mann-Whitney U = 4386, n = 200, $P = 0.128$; Fig. 3, lower

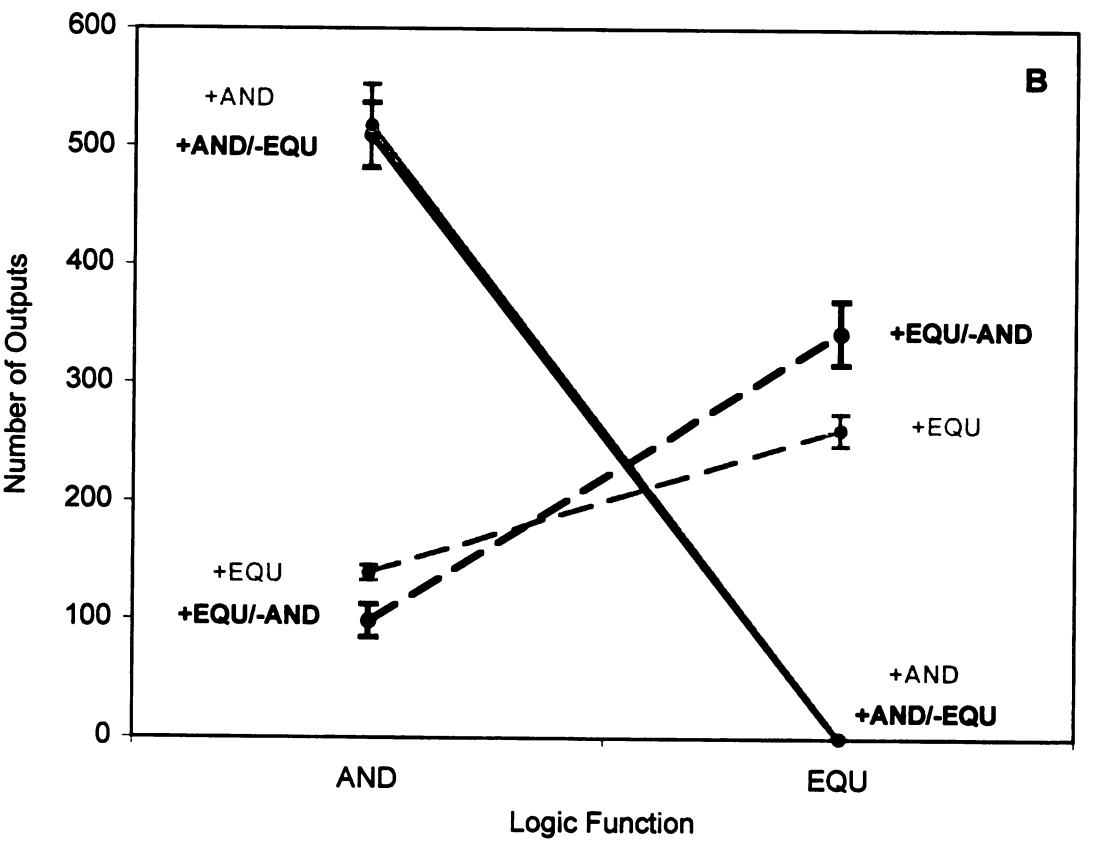
right), although in this case, EQU tended to be performed more often when OR was not punished. EQU was lost when its performance was punished (mean EQU = 0 in the +OR/-EQU environment). However, EQU was also lost when only OR was rewarded, such that its performance declined to zero in the +OR environment. These results reveal an asymmetry in the correlated responses: whereas selection for EQU (+EQU environment) resulted in a correlated increase in the performance of OR, selection for OR (+OR environment) resulted in a complete loss of EQU. Nevertheless, this result is not too surprising in light of the genotype-phenotype map for Ancestor1 (Figure 1). This map shows that all the instructions whose deletions knock out OR also knock out EQU. However, the reverse is not true: not all instructions whose deletions knock out EQU affect the performance of OR. Put another way, the instructions encoding OR are a subset of those encoding EQU. The observed asymmetry in the correlated response thus reflects the underlying asymmetry in the mapping between genotype and phenotype for these two traits.

Direct and correlated responses to selection on functions OR and EQU in Ancestor3

A qualitatively similar pattern emerges when we examine the correlation between OR and EQU in Ancestor3 (Fig. 4A). This ancestor could also perform OR and EQU only once at the start of the experiment. First, selection for EQU (+EQU environment) led to a correlated increase in OR (Fig. 4A, lower left), but selecting on OR (+OR environment) led to the complete loss of EQU (Fig. 4A, lower right). Once again, the direct response to selection was stronger than the correlated response to selection, with the performance of OR higher in the +OR and +OR/-EQU treatments than in the +EQU or

Figure 4. (A) Average number of outputs of OR and EQU in populations evolved from Ancestor3, as a function of their evolution environment. (B) Average number of outputs of AND and EQU.


A
700
600
500
400
300
200
100
0
Number of Outputs
+OR
+OR/-EQU
+EQU/-OR
+EQU
+EQU
+EQU/-OR
+OR
+OR/-EQU
OR
EQU
Logic Function


B
600
500
400
300
200
100
0
Number of Outputs
+AND
+AND/-EQU
+EQU/-AND
+EQU
+EQU
+EQU/-AND
+AND
+AND/-EQU
AND
EQU
Logic Function

+EQU/-OR environments (Fig. 4A; comparing upper left to bottom left). Similarly, the performance of EQU was higher in the +EQU and +EQU/-OR treatments than in the +OR or +OR/-EQU environments (Fig. 4A; comparing upper right to lower right). In both cases, punishing a function seemed to have little effect on the evolution of the selected function, such that the performance of OR did not differ between the +OR and +OR/-EQU treatments (Fig. 4A, upper left). Similarly, the performance of EQU was indistinguishable in the +EQU/-OR and +EQU treatments (Fig. 4A, upper right).

Direct and correlated responses to selection on functions AND and EQU in Ancestor3

At first glance, the pattern looks similar when we examine the functions AND and EQU in populations evolved Ancestor3 (Fig. 4B). Selecting for EQU resulted in a correlated increase in AND from its ancestral level of 1 (Fig. 4B; lower left), but selecting on AND caused EQU to be lost completely (Fig. 4B, lower right). This result is also predicted by the genotype-phenotype map for this ancestor: all instructions that, when deleted, knock out AND also knock out EQU, but the reverse is not true (Fig. 2). In other words, the genome instructions encoding AND are a subset of those encoding EQU.

However, these experiments differ from the preceding ones in two very important respects. First, unlike the other experiments, the performance of AND did not invariably decline to zero when it was selected against (Fig. 4B, lower left). In fact, only 47 of 100 populations evolved in the +EQU/-AND environment actually lost the ability to perform that function (Table 2; far right column). Second, and most significantly, EQU evolved to higher levels when AND was punished than when it was not (average performance of

EQU = 342.8 in +EQU/-AND environment, compared to 261.2 in +EQU environment; Figure 4B, upper right), and this difference was statistically significant (Mann-Whitney U = 5914.5, n = 200, $P = 0.025$).

Table 2. Number of populations (out of a possible 100), that lose a given function, depending on whether it is punished or not. In all experiments, the selected function was rewarded, whereas the other function was not. "Not punished" versus "punished" thus indicates whether this function was simultaneously being punished or not.

Ancestor	Evolution environment: Selected function	Other function	Other function is: Not punished	Punished
Ancestor1	EQU	OR	3	100
	OR	EQU	100	100
Ancestor3	EQU	OR	2	100
	OR	EQU	100	100
	EQU	AND	1	47
	AND	EQU	100	100

One possible explanation is that the increased performance of EQU that evolved in the +EQU/-AND environment did not translate into higher overall fitness. For example, it may have caused a correlated increase in replication rate, such that the two outcomes—although different—represented equally good evolutionary outcomes. To see if this were the case, we took the organisms evolved in the +EQU/-AND environment and transplanted them in the +EQU environment, where we assayed their fitness. Surprisingly, we found that they were also significantly more fit in that environment than the organisms that evolved there (Mann-Whitney U = 6220, n = 200, $P = 0.003$).

The finding that the +EQU/-AND evolved populations, initiated from the same ancestor but evolved in a different environment, have significantly higher fitness in the +EQU environment than the populations that evolved there strongly implies the existence of multiple adaptive peaks. More precisely, it demonstrates that higher fitness is possible in the +EQU environment, and thus, that something prevented the +EQU-evolved populations from reaching that higher fitness. Nevertheless, it should be noted that, although both the increase in EQU and the fitness of populations evolved under the two treatments differ on average, there was substantial variation within each treatment. Thus, the results do not suggest that populations in the +EQU environment always fail to reach the higher fitness peak, but only that they did not reach it as often as populations evolved in the +EQU/-AND environment.

It is unclear what prevented the +EQU-evolved populations from reaching higher fitness. We see at least two related explanations. First, it may be that changing the environment alters the adaptive landscape in such a way that what was previously an adaptive valley becomes flat or uphill (Fig. 5). In this case, the +EQU/-AND populations could have moved into areas of genotypic space that would have been inaccessible in the +EQU environment owing to an intervening adaptive valley. When examined back in the +EQU environment, these +EQU/-AND populations would now sit on or near a higher peak (Fig. 5). This process, whereby selection in a fluctuating environment permits populations to attain a higher fitness than would otherwise be possible, was described by Wright as "mass selection under changing conditions" (Wright 1977).

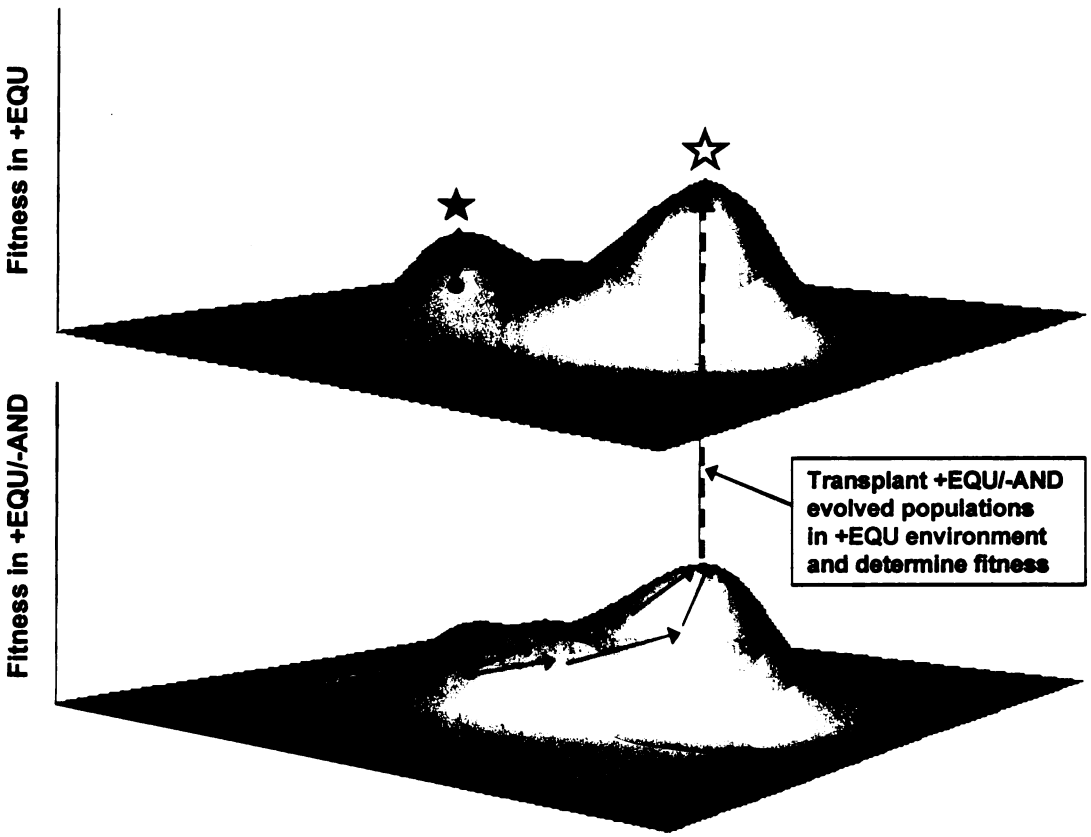


Figure 5. Schematic of a peak-shift in a fluctuating environment. A population initially sits on a peak of relatively low fitness in one environment (top panel). A change in the environment alters the adaptive landscape, such that the intervening valley between that peak and a higher one now becomes uphill, permitting the population to evolve up to the higher peak (lower panel). A subsequent transition back to the original environment results in the population now residing on an alternative adaptive peak.

A second possibility is that the ancestral population sits at some distance from both peaks, and thus either peak can potentially be reached from that starting point. However, in the +EQU environment, selection preferentially moves populations in the direction of the lower peak, either because its ascent is initially steeper, or because there are simply more evolutionary paths that lead to this peak than the other. For example, progress toward the higher peak could involve traversing a single narrow ridge, whereas the path to the lower peak is wide, such that there are many trajectories that lead to the lower peak, but very few that lead to the higher one. The key distinction between this hypothesis and our earlier "Wrightian" one is that there need not be an intervening adaptive valley that prevents the populations from arriving at one peak or another. an intervening adaptive valley that prevents populations from reaching a particular peak. Rather, the evolutionary trajectory of the populations would depend on the likelihood of stumbling upon the rare genetic variants that permit it to travel along the narrow ridge to the higher peak. Moreover, imposing selection against AND in the +EQU/-AND environment would make movement toward a peak that entails the loss of AND more attractive, thereby making the trajectory to this alternative peak more likely in the +EQU/-AND environment.

These hypotheses are not mutually exclusive: the true explanation for the difference in the trajectories of these populations could involve a complex mixture of these processes. Nevertheless, support for the first hypothesis would entail showing that the evolutionary trajectory of populations evolving in the +EQU/-AND environment involved at least some genotypic intermediates that would have been deleterious had they arisen in the

+EQU environment, but instead were neutral or beneficial in the +EQU/-AND environment in which they arose (e.g., Fig. 5). Evidence for the second hypothesis would entail demonstrating that evolution in the +EQU/-AND environment often results in the substitution of the same few mutations in replicate populations, which would support the hypothesis that there was a paucity of paths that lead to this other peak. Below, we present evidence to distinguish between these two possibilities.

Fitness Effects of Mutations that Resulted in the Loss of AND

The only difference between the +EQU and +EQU/-AND environments was that in the latter there was a negative fitness consequence for performing AND. Thus, a good starting place for identifying mutations with differential fitness effects in these environments would be those mutations that caused the loss of AND in the +EQU/-AND environment. In fact, of the 100 populations that evolved in this environment, only 47 of them lost AND (Table 2). This finding alone suggests that the "adaptive valley" explanation is unlikely to explain the results in entirety, unless evolution in the +EQU/-AND environment reduces, but does not completely eliminate, the adaptive valley. Otherwise, we would expect the loss of AND in the +EQU/-AND environment to have occurred more often than it did. Nevertheless, we determined the fitness effect of the mutational steps that caused the loss of AND in these 47 populations. Not surprisingly, they were almost universally beneficial in the +EQU/-AND environment in which they arose (number beneficial = 46, number deleterious = 1). By contrast, 33 of these same 47 steps were deleterious when assayed in the +EQU environment, and a further 10 were neutral, with only 4 being beneficial. This result implies that AND was so often retained

in the +EQU environment at least in part because its loss was consistently associated with a deleterious mutation. Changing the environment by imposing selection against AND thus opened certain evolutionary paths that were not otherwise available. This result thus provides clear support for our first hypothesis, that changing the environment altered the adaptive landscape in such a way that it permitted populations to evolve into regions that would have otherwise been inaccessible and thus to reach a peak of higher fitness.

The Number of Paths Leading to the Loss of AND

Our second hypothesis concerned the relative likelihood of reaching one peak versus another owing to limitations on the production of relevant variation. In this case, we again focus on the mutations that caused the loss of AND. We examined the line of descent in the 47 populations that successfully lost AND when it was selected against and, in each case, we identified the precise genotype along this line of descent that first showed the loss of function. Table 3 shows the alignment of the genome sequences of these genotypes, with the mutations highlighted.

Several patterns are immediately apparent: first, the same few sites are consistently mutated in replicate populations. In multiple instances, populations even converged on the exact same substitution. Second, and even more surprisingly, in 27 of 47 (or 57%) of the cases, the loss of AND was actually caused by a double mutation. Taking into consideration that the genomic mutation rate was 0.1 in these experiments, there should be, on average, only 0.1 mutations per genome per generation. Because genotypes that lie along the line of descent necessarily differ from their parent by at least one mutation,

Table 3. Aligned genome sequences of genotypes along the lines of descent in 47 populations, showing those mutations that caused the loss of AND. The mutations, relative to their parent genotype which could perform AND, are indicated in color: red indicates a point mutations, whereas green and blue indicate insertion and deletion mutations, respectively. In three cases, AND was lost and regained multiple times; we have shown only the first loss of function, and indicated these cases in bold. Owing to their length, several sequences have been trimmed, and thus show only the relevant portion. Images in this thesis are presented in color.

Run ID	Genome Sequence
100	rzavcsqu-axircbngmqqcqa-ppcqppbbcocpqo-g-qlniqfoqfoq----mtttttycstva
142	rzavcwch-mwlvcbooauqcqa-ppcqppbbcocbqo-g-qlnibfoqbyb----abttttycstva
171	rzavcsqp-aaircfogurqcqa-ppcqppbbcocjqo-g-qlniqfoqloq----dtttttycstva
187	rzavcsqx-awircbognzqcqa-ppcqppbbcocsqo-g-qlniqfoqfoq----uttttttycstva
127	rzavcwqx-awircbogmzqcqa-ppcqpprrcoccqo-g-qlniqfoqfoq----uttttttycstva
112	rzavcsqx-apqhcbigqzqcqa-ppcqpphbcocoqo-g-qlniqfoqfoq----uttttttycstva
139	rzavczqx-awircbogmzqcqa-ppcqpphbcocpqo-g-qlnibfoqfoq----ubthctycstva
132	rzavcsqx-awircbogmzqcqa-ppcqpphbcocfqo-g-qlniqfoqfoq----uttttttycstva
119	rzavcsqx-aoircbmqmuqcqa-ppcqppbdcocdqo-g-qlnipfoqfoq----gtttttycstva
117	rzavcsqx-awsicjogmjqcqa-ppcqppbbcocqyo-g-qlnikfoqfoq----uttttttycstva
131	rzavcsqx-awircbogmzqcqa-ppcqppecocqzo-g-qlniqfoqfoq----uttttttycstva
106	rzavcbqx-awircbogizqcqa-ppcqppecocqgo-g-qlniqfoqfaq----uttttttycstva
184	rzavcsqx-awiucbogmzqcqa-ppcqpphcocqgo-g-qlniqfoqfoq----uttttttycstva
108	rzavcsqx-dwircbmgmuqcqa-ppcqpprcocqto-g-qlniqdoqfoq----mtqtttycstva
148	rzavcsqxtawkvcbbgmzqcqa-ppcqppdcocqdo-g-qlniqfoqfoq----yvttttycstva
101	rzavcszx-bwrkzx-gmuqcqa-ppcqppqcocqqo-gjqlnidfoqfaq----mtjctfycstva
121	rzavcuqx-awirxefgizqcqa-ppcqppqcocqqo-gnqlnibfoaboe----btttytycstva
175	rzavcszb-awirxbwgxzqcqa-ppcqppqcocqqo--aqlniiaoqfoq----btotttycstva
164	rzavcyiawwihbbzgemuqcqa-ppcqppbcocqdo-g-qlniqfoqnoyb----ttbttycstva
144	rzavcsqx-awircbogmzqcqc-ppcqppvcocqqo-g-qlniqfoqfoqu----tttttycstva
116	rzavcsqx-zwircboxmzqcqa-ppcqpp*cocqqo-g-qlniqfoqfoq----dtttttycstva
143	gzavcphc-uwwiiuwayuqcqa-ppcqppqcocqqo-g-qlniqfoqdoyb----cofdfycstva
135	rzavcgeo-awiqfbocmuqcqatppcqppqcocqqo-glqlniqfoqfoq----ytttttycstva

Table 3. (continued)

Run ID	Genome Sequence
104	rzavcsqm-kwhcxlgmuq-cqa-ppcqppqcocqqogglqltniqfoqfoq---ytttttycstva
185	rzavcsqh-gnirqblgguqcqa-ppcqppqcocqqo-gjqlniqfoaobq----dtttytycstva
186	rzavcsqx-rbiyksnomuqcqa-ppcqppqcocqqo-gjqlnibfoacoq----btttytycstva
137	rzavcfbm-ewisebogmuqcqa-ppcqppqcocqqo-gjqlniofogfaq----qbytetycstva
192	rzavcsdx-awircbogozqcqa-ppcqppqcocqqo-gjqlniqfoqf*q----uttttttycstva
153	rzavc*qx-alorrbogauqcqa-ppcqppqcocqqo-gjqlniqfoaboq----rtttttycstva
154	izazcqdh-mguiqnugcqqcqa-ppcqppqcocqqo-gjqlncqfogfcq----bfbyttycstvs
118	rzavcfqz-dwircbogmuqcqa-ppcqppqcocqqo-gkqlnibfoarog----qevtttycstva
176	rzavcsqz-awbr*rogmuqcqa-ppcqppqcocqqo-gkqlniqfoqfbq----bttttttycstva
178	rzavcjyj-mywxcshaiuqcqa-ppcqppqcocqqo-gkqlnibfoqfoc----jtttytycstva
138	rzavchqf-cwirfb-gquqcqa-ppcqppqcocqqo-glqlnibfogqfh----xttqytycstva
179	rzavckqc-gxlryihgmuqcqa-ppcqppqcocqqo-glqlniqfoqfbq----ettttytycstva
162	rzavcsqj-dpsrzzofcuqcqa-ppcqppqcocqqo-glqlniqfoaboq----f*ttttycstva
190	rzavcjqa-qxiriioymuqcqa-ppcqppqcocqqo-gmqlniqfogfaq----x*ttttycstva
188	ahvvcsuj-aeikcuoxxzqcqa-ppcqppqcocqqo-gmqlnibfoaboq----ftttytycstva
124	rzavcgqj-j*irxnogauqcqa-ppcqppqcocqqo-gmqlniqbobfoq----xttttttycstva
129	rzavctfc-wafzczbdafqcqa-pocqppqcocqqo-gnqlnibfgotqtfotquatmttycstva
107	rzavc*wo-rwxrabogcuqcqa-ppcqppqcocqqo-gnqlncbfogfax----qytettycstva
147	rzavcpqn-awirlifxguqcqa-ppcqppqcocqqo-goqlnibfoqyba----uybtdttycsva
130	rzavcggd-awrecbngkuqcqa-ppcqppqcocqqo-goqlnibfobfcg----qytewtycstva
134	rzavcojx-ahircqdaxzqcqa-ppcqppqcocqqo-goqltniqfoqfh---qytttttycstva
168	ahvcsmlh-fyibcmolmuqcqa-ppcqppqcocqqo-gpqlncqfogfaq-----ocyttycs-va
102	rzfrzavcqqflwipcbobmzqcqappcqppqcocqqogmqlncqbkmfoq----ubttytycstib
123	rzavcsdx-rwircbogmuqcqa-ppcqppqcocqqo-gpqlniqfoqfoq----qtrtttycsqva

we calculated the probability that a particular genotype would differ from its parent genotype by two or more mutations. Performing this calculation, we find that roughly only 0.048 or 4.8% of the genotypes that differ by at least one mutation are expected to differ by two or more. Thus, the mutations causing the loss of AND resulted from double mutations nearly 12 times more often than expected by chance. This result strongly suggests that there were a very limited number of ways to produce the desired phenotype—which in this case, meant eliminating AND while at the same time retaining EQU and other aspects of organismal performance. Importantly, we know that there are many ways simply to lose AND, as evidenced by the genotype-phenotype map (Fig. 2), which shows some of the mutations that knock out this function. Most likely, it is the pleiotropic effect that losing AND usually has on EQU or other fitness components that places such severe limitations on the particular mutations that can be substituted in the +EQU/-AND environment.

Finally, the overrepresentation of double mutations among those causing the loss of AND is interesting its own right, because it implies that the component mutations were not beneficial when they arose individually in the +EQU/-AND environment. It therefore suggests the presence of an intervening adaptive valley in the +EQU/-AND environment, albeit a narrow one that could be traversed by a double mutation. This result also may explain why so few populations evolved in this environment lost the AND function, despite selection in favor of its loss. Although selecting against AND did cause its loss to occur far more often than not selecting against it (Table 2), it appears that the mutations required to produce this loss were so severely limited that, even in this penalizing

environment, many populations still failed to lose it. Thus, our results demonstrate that constraints on the types of variation available to these populations may have also limited their ability to reach the alternative, higher adaptive peak.

Discussion

Digital organisms are self-replicating computer programs that mutate, adapt, and evolve. They also possess a complex mapping between genotype and phenotype, a property that gives rise to pleiotropy and epistasis. We examined their importance for the adaptive dynamics of digital organisms and, in particular, the role of pleiotropy in constraining the adaptation of one function owing to its genetic association with another. Our results revealed that correlated responses were often highly asymmetric, but that this asymmetry was consistent with the way in which these functions are encoded in the genome. Both ancestors' genomes showed that in many cases, the sites that encoded one function were necessary for the other function, but that the overlap was not complete. Rather, the instruction encoding OR or AND were often a subset of those required for the EQU function. This asymmetry in the encoding of these functions is thus reflected in the asymmetry of their correlated responses to selection.

Our results also show that, in most cases, functions experienced greater increases in performance in response to direct selection rather than correlated selection on another function. In the most extreme example, the function EQU was lost in every environment in which it was not directly selected (Table 2). This result is consistent with previous

work demonstrating that the EQU function fulfills the definition of a complex feature (Lenski et al. 2003a). Because its evolutionary emergence requires building upon simpler functions, it makes sense that these functions will usually comprise a subset of the instructions that encode EQU. Moreover, the failure to be robust to mutations is a diagnostic feature of these traits; its loss in the absence of direct selection to maintain it is therefore not surprising.

When we consider the functions that show correlated increases in response to selection for EQU, we also find that these functions are usually lost when selected against, illustrating that even apparently strongly coupled traits can be disassociated. Other studies that have examined selection on correlated traits have shown that these associations can be modified by selection. For example, Lenski (1988) show that the cost associated with a resistance allele could be reduced through further substitution of modifier alleles. Similarly, Zijlstra et al. (2003) found that selection to uncouple a correlation between development time and eyespot size in butterflies was not only possible, but that the response to selection in lines selected in the uncoupled direction was also faster than expected based on quantitative genetic estimates. Thus, even short-term responses can differ from expectations based on current genetic variation. In this system, the presence or absence of a correlation between traits may reflect the nature of past selection better than it predicts the response to future selection. This idea is analogous to that proposed by Wright to explain the maintenance of vestigial traits, who commented that useless or even slightly deleterious parts may be retained for long periods of time owing to their pleiotropic relationship to characters under positive

selection for their retention. He emphasized, however, that these parts might be rapidly lost during times of "reorganization" (Lande 1978; Wright 1984).

While nearly all of the correlations could be consistently broken when one of the functions was punished, there was one interesting and striking exception to this pattern (Table 2). In this case, over half of the 100 replicate populations evolved in environments where EQU was selected for increases but AND was selected against failed to lose the latter function. Even more surprising, populations evolved in the environment where AND was selected against evolved higher performances of EQU than populations evolving in environments where AND was not selected against. This difference in the performance of EQU also translated into higher overall fitness, such that populations evolved in the +EQU/-AND environment were significantly more fit in the +EQU environment than the populations that evolved in it. This finding strongly indicated the presence of multiple adaptive peaks in the +EQU environment. More important, it implies that populations evolving in the +EQU environment were somehow prevented from reaching this peak of higher fitness.

We considered two different hypotheses to explain the failure of +EQU-evolved populations to reach the higher fitness peak that was achieved by populations that evolved in the +EQU/-AND environment. The first hypothesis involved a shift in the adaptive landscape between the +EQU/-AND and the +EQU environments, which allowed populations to cross what had previously been an adaptive valley and thus to reach a peak of higher fitness. Consistent with this interpretation, mutations causing the

loss of AND were nearly always beneficial in the +EQU/-AND environment, but these same mutations would have been deleterious had they arisen in the +EQU environment. The finding that these mutations were usually deleterious in the +EQU environment helps to explain why so few populations evolved in that environment lost the AND function. In other words, the evolutionary trajectories of populations evolving in the +EQU/-AND environment often progressed through genotypic intermediates that would not have been selectively favored in the +EQU environment, and thus, the populations traversed regions of the fitness landscape that were inaccessible to populations evolving the +EQU environment.

We also hypothesized a second explanation for the failure to find this alternative adaptive peak. This alternative process suggests that the populations start at some distance from both peaks. The genetic coupling of traits (in this case, of EQU and AND) results in a bias in the available genetic variation and thereby predisposes the population to evolve towards one of the peaks. The importance of correlated traits for movement on rugged adaptive landscapes has been considered in detail by Price et al. (1993), who describe how selection on a correlated trait can cause a population to shift between two alternative adaptive peaks for some focal trait. Similarly, Schluter (1996; 2000) illustrated how a genetic correlation could lead to evolution along "genetic lines of least resistance." The larger importance of these models is that they illustrate how limits to the direction of genetic variation can alter the evolutionary trajectory of a population evolving on a rugged adaptive landscape.

Our results indicate that most of the mutations causing AND to be lost were subject to severe constraints owing to their pleiotropic effects on EQU or some other component of fitness. Closer examination of these mutations revealed that, in many cases, they increased the generation time of the organism and thus were unlikely to be fixed except in environments where AND was being punished. In such cases, the benefit conferred by the loss of AND offset the other fitness costs of the mutations, permitting them to fix, but only in the +EQU/-AND environment. This result is particularly interesting in light of the finding that the mutations causing AND to be lost occurred at the same handful of sites in the genome. In many cases, these losses were caused by double mutations. Thus, it appears that the mutations that fix were limited to those that conferred a net benefit in the +EQU/-AND environment, and that the paucity of such mutations in the genetic background of Ancestor3 is what led to the high degree of parallelism in the substitutions (Table 3). An interesting corollary to this result is that if the fitness penalty for performing AND were made larger, and thus selection for its loss were stronger, then a wider range of mutations might confer a net benefit in the +EQU/-AND environment, despite their deleterious pleiotropic effects. In other words, stronger selection against AND is expected to produce a more diverse set of substitutions associated with its loss. This hypothesis is somewhat counterintuitive, but is consistent with theory regarding the fixation of major mutations with deleterious pleiotropic effects, as it is a process that is favored by strong selection (Lande 1983). An interesting follow-up to these experiments would thus be to examine the patterns of genomic evolution in this system as a function of the magnitude of the selection coefficient against AND.

Epistasis for fitness is required to generate rugged adaptive landscapes (Brodie 2000; Whitlock et al. 1995; Wright 1984). The work here and elsewhere (Lenski et al. 2003a) indicates that such landscapes emerge even in simple digital systems. While there is considerable evidence that rugged adaptive landscapes also exist in nature, Wright's Shifting Balance theory, which incorporates a particular set of evolutionary forces on these landscapes to explain adaptation, remains controversial to this day (Coyne et al. 1997; Coyne et al. 2000). Part of the difficulty is that it is nearly impossible to determine in retrospect whether peak shifts have occurred and, if so, by what mechanism. We interpret our results as an indication of how universal rugged adaptive landscapes may be, but we also note that the peak shift we observed was not the result of drift driving the population through an adaptive valley. Quite the opposite: 99 of 100 populations evolved in the +EQU environment never found the alternative adaptive peak. Only by changing the environment, such that movement toward the alternative peak became more strongly favored, were we able to observe the occurrence of a peak shift. Thus, our results do not provide support for Shifting Balance Theory, although they do provide evidence for Wright's alternative hypothesis that mass selection in a changing environment can allow a peak shift to occur.

In closing, digital systems offer an excellent opportunity to explore complex evolutionary dynamics. In addition to the difficulties faced by natural systems, theoretical analyses have been hampered by the complexity of the landscapes, which usually limits their scope to considerations of one or a few loci or else requires other simplifying assumptions. Digital systems permit experiments on adaptation in highly

complex and multi-dimensional landscapes, and thus offer the opportunity to develop and test theories regarding the causes and consequences of evolution on rugged adaptive landscapes in ways that have not previously been possible.

APPENDICES

APPENDIX A. Supplemental Data from Ch.1

Table A1. Results of one-way ANOVAs testing the effect of genotype in each novel resource.

Resource	df	MS(Genotype)	MS(Error)	*F*	*P*
NAG	26, 54	0.0040	0.0027	1.49	0.1098
Mannitol	26, 54	0.0042	0.0022	1.88	0.0258
Maltose	26, 54	0.0044	0.0021	2.08	0.0119
Galactose	26, 53	0.0072	0.0021	3.46	<0.0001
Melibiose	26, 54	0.0427	0.0078	5.47	<0.0001

Table A2. Results of one-way ANOVAs for the fitness effect of each genotype across resources.

Genotype	df	MS(Resource)	MS(Error)	*F*	*P*
9962	4, 9	0.0097	0.0013	7.38	0.0064
9968	4, 10	0.0041	0.0023	1.80	0.2064
9970	4, 10	0.0092	0.0019	4.84	0.0197
9972	4, 10	0.0491	0.0058	8.44	0.0030
9974	4, 10	0.0300	0.0040	7.44	0.0048
9976	4, 10	0.0053	0.0041	1.29	0.3363
9978	4, 10	0.0033	0.0040	0.83	0.5369
9980	4, 10	0.0013	0.0020	0.63	0.6497
9982	4, 10	0.1013	0.0062	16.25	0.0002
9984	4, 10	0.0049	0.0057	0.87	0.5142
9986	4, 10	0.0048	0.0018	2.61	0.0997
9988	4, 10	0.0033	0.0022	1.47	0.2821
9990	4, 10	0.0185	0.0042	4.47	0.0250
9992	4, 10	0.0205	0.0023	9.02	0.0024
9994	4, 10	0.0736	0.0030	24.24	<0.0001
9996	4, 10	0.0087	0.0020	4.39	0.0262
9998	4, 10	0.0592	0.0104	5.70	0.0118
10000	4, 10	0.0077	0.0016	4.95	0.0183
10002	4, 10	0.0044	0.0025	1.74	0.2185
10004	4, 10	0.0075	0.0016	4.59	0.0232
10006	4, 10	0.1205	0.0040	30.01	<0.0001
10008	4, 10	0.0060	0.0020	3.08	0.0677
10012	4, 10	0.0145	0.0036	4.07	0.0328
10014	4, 10	0.0118	0.0020	5.79	0.0112
10016	4, 10	0.0438	0.0034	12.85	0.0006
10018	4, 10	0.0632	0.0056	11.23	0.0010
10020	4, 10	0.0238	0.0018	13.05	0.0006

APPENDIX B. Supplemental Information on the Sequenced Loci

spoT – The *spoT* locus is responsible for both the synthesis and degradation of a molecule known as guanosine tetraphosphate, or ppGpp. ppGpp is associated with the stringent response, which is the physiological response of bacterial cells to nitrogen or carbon starvation (Cashel et al. 1996). The production of ppGpp inhibits stable RNA synthesis, which in turn results in a decrease in nearly all metabolic activities of the cell, including transcription, translation, and DNA replication. Increases in ppGpp are also thought to result in increased transcription of stationary phase genes, which improve survival during times of low resources. Mutational studies of the *spoT* locus indicate that the regions required for the synthesis and degradation of ppGpp are partly overlapping (Gentry and Cashel 1996). There is also a third region of unknown function.

nadR – The *nadR* gene encodes a DNA binding protein that acts as a repressor of the nicotinamide adenine dinucleotide (NAD) biosynthesis genes. It is known to have a secondary role in the transport of nicotinamide mononucleotide (NMN) into the cell. (NAD can be broken down into NMN and AMP, such that the uptake of NMN amounts to the scavenging of NAD precursors.) Evidence suggests that the 5' end of the gene may be necessary for the repressor role, whereas the 3' end appears to be necessary for NMN transport (Penfound and Foster 1996). Mutations in the central portion of the gene have been shown to produce a super-repressor phenotype, with enhanced repressor functions but reduced transport functions.

pykF – The *pykF* locus encodes pyruvate kinase I, an enzyme in the glycolysis pathway that converts phosphoenolpyrute (PEP) and ADP to pyruvate and ATP. It is known that the PEP:pyruvate ratio is an important factor determining the phosphorylation state of the PTS enzyme IIA^{Glc}, which aids in the transport of glucose across the inner membrane and can also bind to enzymes involved in the metabolism of non-PTS resources (Hogema et al. 1998).

pbp-rodA - The *pbp-rodA* encodes two genes, *pbp* which encodes a penicillin-binding protein called PBP2, and *rodA*, which encodes an integral membrane protein required for the proper activity of PBP2. Together, they are thought to determine the rod shape of the bacterial cell (Lutkenhaus and Mukherjee 1996).

hokB-sokB – The *hokB-sokB* locus encodes a toxin-antitoxin pair that are homologous to those associated with plasmid stability (Pedersen and Gerdes 1999). Hok stands for "host-killing", whereas sok stands for "suppression of killing". These genes may function in programmed cell death.

Rbs operon – The ribose operon is required for the catabolism of ribose. It contains six genes (*rbsDACBKR*). Previous work in the long-term lines showed losses of the ability to catabolize ribose, which corresponded to large deletions in the ribose operon that ranged in size from a deletion of the promoter, *rbsD* and part of *rbsA*, to the entire operon and a portion of a neighboring gene (Cooper et al. 2001).

APPENDIX C. Locations of identified mutations

Genotype	Gene	Position	Mutation	Amino Acid Change
9962	*spoT*	1715	A → C	Lys → Thr
9968	*spoT*	1993	C → T	Arg → Cys
9970	*spoT*	1769	A → C	Lys → Thr
9974	*spoT*	316	C → T	Leu → Phe
9976	*spoT*	1324	A → C	Thr → Pro
9980	*spoT*	1724	G → T	Arg → Leu
9984	*spoT*	1226	T → C	Phe → Ser
9986	*spoT*	1370	G →T	Trp → Leu
9988	*spoT*	1324	A → C	Thr → Pro
9990	*spoT*	1994	G → A	Arg → His
9996	*spoT*	990	G → A	Met → Ile
10000	*spoT*	1249	A → C	Ile → Leu
10002	*spoT*	1249	A → C	Ile → Leu
9982	*nadR*	30	A → Δ[1]	deletion
9992	*nadR*	169	::IS *150*	insertion
10004	*nadR*	931	A → G	Lys → Glu
10014	*nadR*	186 – 189[2]	G → Δ	deletion
10020	*nadR*	186 – 189[2]	G → Δ	deletion
9998	*hok-sok*		::IS *150*	insertion
10006	*pbp-rodA*	-828[3]	C → A	noncoding
9990	*pykF*	1153	C → A	Arg → Ser

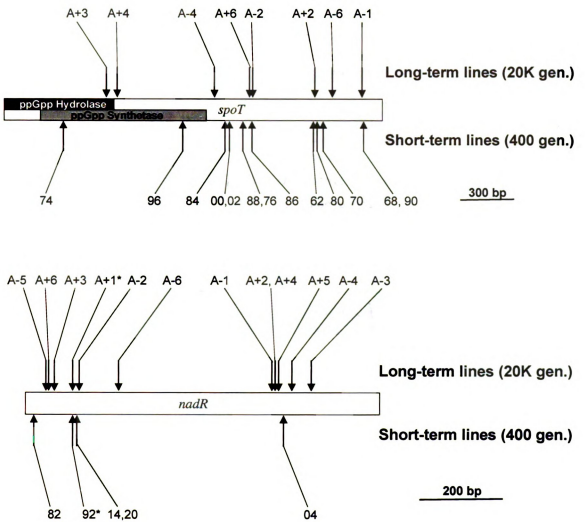
Appendix C. Location of all identified mutations. Double-mutants are listed twice, and their genotype identities are in boldface. The position is relative to that of the first basepair of the listed gene's start codon.

[1]Δ indicates a deletion mutation

[2]Deletion of one G in a string of 4 G's.

[3]Number is negative to indicate the number of basepairs upstream from the start of the *pbp-rodA* genes.

APPENDIX D. Locations of identified mutations in *spoT* and *nadR*.



Appendix D. (A) Location of mutations found in the *spoT* locus after 400 generations (bottom; this study) and 20,000 generations (top; Cooper et al. 2003). Genotype identification numbers listed in Appendix C have been abbreviated here to include only their last two digi ts, which are unique to each genotype (e.g., 9990 is listed as 90). The regions corresponding to the ppGpp hydrolase and synthetase have been shown (Gentry and Cashel 1996). Arrows indicate locations along the gene, from the N-terminus (left) to the C-terminus, positioned according to the number in the amino acid sequence, such that mutations that affect the same amino acid site are indicated by a single arrow. (B) Location of mutations found in the *nadR* locus after 400 generations (bottom; this study) and 20,000 generations (top; Woods et al., manuscript). Asterisks indicate mutations caused by the insertion of IS*150*.

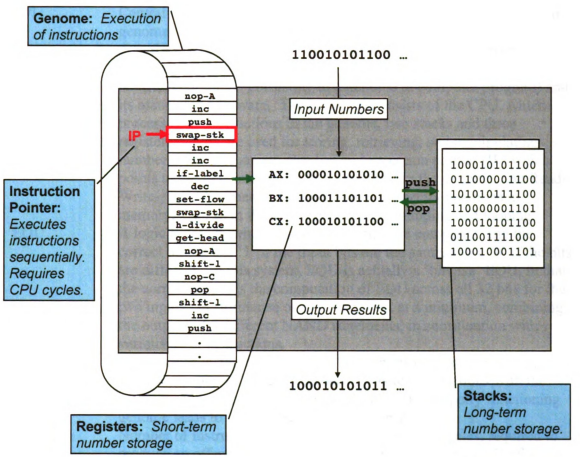


Figure E1. Schematic of a digital organism in *Avida*. A digital organism consists of a genome (computer program), three registers, two stacks, and four heads (one of which is the Instruction Pointer). Execution of the program requires CPU cycles, and the point of execution is indicated by the location of the instruction pointer (IP). An input-output (I/O) instruction enables an organism to input binary numbers into the registers and output results of computations. Most of the instructions in the genome operate directly on the numbers in the register, although the `push` and `pop` instructions cause numbers in the registers to pushed onto the stack or popped off of the stack, respectively. The stacks are thus primarily used for storing numbers, whereas the registers are used to manipulate them. Images in this thesis are presented in color.

Table E1: Glossary of Terms

Terms	Definition
CPU	Central Processing Unit. All organisms have the instructions in their genomes executed by a virtual CPU. A mutation that causes an organism to have more CPU cycles (that is, to have its genome executed faster than others) is generally beneficial.
Digital Organism	A virtual computer, consisting of a genome (a computer program) and its associated hardware. The hardware consists of the CPU, which processes the instructions in the genome, two stacks and three registers, which are used for storing, retrieving, and manipulating numbers. Each organism also has an instruction pointer (IP) which points to the next instruction to be executed in the genome, and Read-, Write-, and Flow-heads, which are used to specify positions in memory, such as in the copy process or for jumping and looping.
EQU	A logic function, where two binary inputs are compared, and the correct output is a '1' if the input bits are the same, and a '0' if the bits are different. In this system, EQU is actually a 'bitwise' EQU, in that the correct output is the computation of EQU across all 32 bits for the two inputs. Performance of EQU requires, at a minimum, combining the outputs of 5 different NAND statements, in coordination with various other instructions.
Genome	Sequence of instructions that may contain information for making duplicate copies of the genome, as well as for interactions with the environment. Execution of the instructions in a properly functioning genome leads to the production of an offspring.
Gestation Time	Number of instructions executed, and hence CPU cycles, required to produce an offspring. Gestation time is generally a multiple of genome length, but varies as a function of the efficiency of the copy process and the number of loops in execution.
Instruction	Units that comprise the genome. Each site in the genome is 1 of 26 possible instructions. Instructions not present in an ancestral genome may be introduced into the genome of descendents via mutation.
Logic Function	Computations based on binary inputs. Organisms may evolve to perform bitwise logic functions based on numbers they input from the environment.
Mutations	Mutations can be point mutations, where one instruction is randomly replaced with another during the copy process. Mutations can also be insertion or deletion mutations, causing genome size to grow or shrink in length. The rates of point, insertion, and deletion mutations are specified by the experimenter.
NAND	One of the 26 possible instructions in the genome. Also a core logic function; all other logic functions can be built from combinations of NANDs.

APPENDIX F: Table of Punishments and Rewards

Table F1. Punishments and rewards were used in the calculation of merit, which is directly proportional to the number of CPU cycles an organism receives. Rewards were proportional to the difficulty of the function (i.e., EQU is the most difficult), and were constant for a particular function regardless of the particular treatment. Punishments were less than or equal to the magnitude of the reward, to ensure that the total bonus was not less than zero.

Environment	Punished Function	Bonus	Rewarded Function	Bonus
+EQU	-----	-----	EQU	+5
+AND	-----	-----	AND	+2
+OR	-----	-----	OR	+3
+EQU/-AND	AND	-2	EQU	+5
+AND/-EQU	EQU	-2	AND	+2
+EQU/-OR	OR	-3	EQU	+5
+OR/-EQU	EQU	-3	OR	+3

LITERATURE CITED

AGRAWAL, A. A. 2000. Host range evolution: adaptation and tradeoffs in fitness of mites on alternative hosts. Ecology 81:500-508.

ANDERSSON, S. G. E., A. ZOMORODIPOUR, J. O. ANDERSSON, et al. 1998. The genome sequence of *Rickettsia prowazekii*. Nature 396:133-143.

BARRIER, M., R. H. ROBICHAUX, and M. D. PURUGGANAN. 2001. Accelerated regulatory gene evolution in an adaptive radiation. Proc. Natl. Acad. Sci. USA 98:10208-10213.

BARTON, N., and L. PARTRIDGE. 2000. Limits to natural selection. Bioessays 22:1075-1084.

BARTON, N. H., and P. D. KEIGHTLEY. 2002. Understanding quantitative genetic variation. Nat. Rev. Genet. 3:11-21.

BELDADE, P., K. KOOPS, and P. M. BRAKEFIELD. 2002. Developmental constraints versus flexibility in morphological evolution. Nature 416:844-847.

BOROWSKY, R., and H. WILKENS. 2002. Mapping a cave fish genome: Polygenic systems and regressive evolution. J. Hered. 93:19-21.

BRODIE, E. D. I. 2000. Why evolutionary genetics does not always add up. Pp. 3-19 *in* J. B. Wolf, E. D. I. Brodie and M. J. Wade, eds. Epistasis and the Evolutionary Process. Oxford University Press, Oxford.

CASHEL, M., D. R. GENTRY, V. J. HERNANDEZ, et al. 1996. The stringent response. Pp. 1458-1496 *in* F. C. Neidhardt, ed. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. ASM Press, Washington, D.C.

CHEN, L., A. L. DEVRIES, and C.-H. C. CHENG. 1997. Evolution of antifreeze glycoprotein gene a trypsinogen gene in Antarctic notothenioid fish. Proc. Natl. Acad. Sci. USA 94:3811-3816.

CLARKE, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18:117-143.

CLARKE, K. R., and R. N. GORLEY. 2001. PRIMER v5: User manual/tutorial. PRIMER-E, Plymouth, UK.

COLE, S. T., K. EIGLMEIER, J. PARKHILL, et al. 2001. Massive gene decay in the leprosy bacillus. Nature 409:1007-1011.

CONNER, J. K. 2003. Artificial selection: A powerful tool for ecologists. Ecology 84:1650-1660.

COOPER, T. F., D. E. ROZEN, and R. E. LENSKI. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 100:1072-1077.

COOPER, V. S., and R. E. LENSKI. 2000. The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature 407:736-739.

COOPER, V. S., D. SCHNEIDER, M. BLOT, et al. 2001. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli B*. J. Bacteriol. 183:2834-2841.

COYNE, J. A., N. H. BARTON, and M. TURELLI. 1997. Perspective: A critique of Sewall Wright's shifting balance theory of evolution. Evolution 51:643-671.

---. 2000. Is Wright's shifting balance process important in evolution? Evolution 54:306-317.

CRESPI, B. J. 2000. The evolution of maladaptation. Heredity 84:623-629.

CROZAT, E., N. PHILIPPE, R. E. LENSKI, et al. 2005. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. Genetics 169:523-532.

DARWIN, C. 1859. On the Origin of Species by Means of Natural Selection. John Murray, London.

DE VISSER, J. A. G. M., C. W. ZEYL, P. J. GERRISH, et al. 1999. Diminishing returns from mutation supply rate in asexual populations. Science 283:404-406.

DEAN, A. M., and G. B. GOLDING. 1997. Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase. Proc. Natl. Acad. Sci. USA 94:3104-3109.

DOBZHANSKY, T. 1937. Genetics and the Origins of Species. Columbia University Press, New York.

ELENA, S. F., V. S. COOPER, and R. E. LENSKI. 1996. Punctuated evolution caused by selection of rare beneficial mutations. Science 272:1802-1804.

FALCONER, D. S., and T. F. C. MACKAY. 1996. Introduction to Quantitative Genetics. Harlow, Longman.

FISHER, R. A. 1958. The genetical theory of natural selection. Dover, New York.

FRY, J. D. 1993. The general vigor problem: Can antagonistic pleiotropy be detected when genetic covariances are positive? Evolution 47:327-333.

---. 1996. The evolution of host specialization: are trade-offs overrated? Am. Nat. 148:S84-S107.

---. 2003. Detecting ecological trade-offs using selection experiments. Ecology 84:1672-1678.

FUTUYMA, D. J., and G. MORENO. 1988. The evolution of ecological specialization. Annu. Rev. Ecol. Syst. 19:207-233.

GAVRILETS, S., and A. HASTINGS. 1996. Founder effect speciation: a theoretical reassessment. Am. Nat. 147:466-491.

GENTRY, D. R., and M. CASHEL. 1996. Mutational analysis of the *Escherichia coli spoT* gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation. Molecular Microbiology 19:1373-1384.

GERRISH, P. J., and R. E. LENSKI. 1998. The fate of competing beneficial mutations in an asexual population. Genetica 103:127-144.

GOLDBERG, D. E., and S. M. SCHEINER. 2001. ANOVA and ANCOVA: Field competition experiments. Pp. 77-98 *in* S. M. Scheiner and J. Gurevitch, eds. Design and analysis of ecological experiments. Oxford University Press, New York.

HOGEMA, B. M., J. C. ARENTS, R. BADER, et al. 1998. Induced exclusion in *Escherichia coli* by non-PTS substrates: the role of the PEP to pyruvate ratio in determining the phosphorylation state of enzyme IIA^{Glc}. Molecular Microbiology 30:487-498.

JACOB, F. 1977. Evolution and tinkering. Science 196:1161-1166.

JAENIKE, J. 1990. Host specialization in phytophagous insects. Annu. Rev. Ecol. Syst. 21:243-273.

JEFFREY, W. R., STRICKLER, A.G., AAND YAMAMOTO, Y. 2003. To see or not to see: Evolution of eye degeneration in Mexican blind cavefish. Integrative Comparative Biology 43:531-541.

JERNIGAN, R. W., CULVER, D.C., FONG, D.W. 1994. The dual role of selection and evolutionary history as reflected in genetic correlations. Evolution 48:587-596.

JOHNSON, T., and N. H. BARTON. 2002. The effect of deleterious alleles on adaptation in asexual populations. Genetics 162:395-411.

JONES, C. D. 1998. The genetic basis of *Drosophila sechellia*'s resistance to a host plant toxin. Genetics 149:1899-1908.

JONES, R., CULVER, D.C., KANE, T.C. 1992. Are parallel morphologies of cave organisms the result of similar selection pressures? Evolution 46:353-365.

KASSEN, R. 2002. The experimental evolution of specialists, generalists, and the maintenance of diversity. J. Evol. Biol. 15:173-190.

KAWECKI, T. J. 1994. Accumulation of deleterious mutations and the evolutionary cost of being a generalist. Am. Nat. 144:833-838.

---. 1998. Red Queen meets Santa Rosalia: Arms races and the evolution of host specialization in organisms with parasitic lifestyles. Am. Nat. 152:635-651.

---. 2000. The evolution of genetic canalization under fluctuating selection. Evolution 54:1-12.

KING, M. C., and A. C. WILSON. 1975. Evolution at two levels in humans and chimpanzees. Science 188:107-116.

KORONA, R., C. H. NAKATSU, L. J. FORNEY, et al. 1994. Evidence for multiple adaptive peaks from populations of bacteria evolving in a structured habitat. Proc. Natl. Acad. Sci. USA 91:9037-9041.

LANDE, R. 1978. Evolutionary mechanisms of limb loss in tetrapods. Evolution 32:73-92.

---. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33:402-416.

---. 1983. The response to selection on major and minor mutations affecting a metrical trait. Heredity 50:47-65.

LANDE, R., and S. J. ARNOLD. 1983. The measurement of selection on correlated characters. Evolution 37:1210-1226.

LENSKI, R. E. 1988. Experimental studies of pleiotropy and epistasis in *Escherichia coli*. 2. Compensation for maladaptive effects associated with resistance to virus T4. Evolution 42:433-440.

LENSKI, R. E., C. OFRIA, R. T. PENNOCK, et al. 2003a. The evolutionary origin of complex features. Nature 423:139-144.

LENSKI, R. E., M. R. ROSE, S. C. SIMPSON, et al. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am. Nat. 138:1315-1341.

LENSKI, R. E., C. L. WINKWORTH, and M. A. RILEY. 2003b. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. J. Mol. Evol. 56:498-508.

LEVINS, R. 1968. Evolution in Changing Environments. Princeton University Press, Princeton.

LUTKENHAUS, J., and A. MUKHERJEE. 1996. Cell division. Pp. 1615-1639 *in* F. C. Neidhardt, ed. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. ASM Press, Washington, D.C.

MACLEAN, R. C., G. BELL, and P. B. RAINEY. 2004. The evolution of a pleiotropic fitness tradeoff in *Pseudomonas fluorescens*. Proc. Natl. Acad. Sci. USA 101:8072-8077.

MANI, G. S., and B. C. CLARKE. 1990. Mutational order: a major stochastic process in evolution. Proc. R. Soc. Lond. B 240:29-37.

MELÉNDEZ-HEVIA, E., T. G. WADDELL, and M. CASCANTE. 1996. The puzzle of the Krebs citric acid cycle: assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution. Journal of Molecular Evolution 43:293-303.

MULLER, H. J. 1932. Some genetic aspects of sex. Am. Nat. 66:118-138.

NACHMAN, M. W. 2005. The genetic basis of adaptation: lessons from concealing coloration in pocket mice. Genetica 123:125-136.

NILSSON, D.-E., and S. A. PELGER. 1994. A pessimistic estimate of the time required for an eye to evolve. Proc. R. Soc. Lond. B 256:53-58.

OCHMAN, H., and N. A. MORAN. 2001. Genes lost and genes found: Evolution of bacterial pathogenesis and symbiosis. Science 292:1096-1098.

OFRIA, C., and C. O. WILKE. 2004. Avida: A software platform for research in computational evolutionary biology. Artificial Life 10

ORR, H. A. 2000. The rate of adaptation in asexuals. Genetics 155:961-968.

ORR, H. A., and J. A. COYNE. 1992. The genetics of adaptation: a reassessment. Am. Nat. 140:725-742.

OTTO, S. P. 2004. Two steps forward, one step back: the pleiotropic effects of favored alleles. Proc. R. Soc. Lond. B 271:705-714.

PECK, J. R. 1994. A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. Genetics 137:597-606.

PEDERSEN, K., and K. GERDES. 1999. Multiple *hok* genes on the chromosome of *Escherichia coli*. Molecular Microbiology 32:1090-1102.

PENFOUND, T., and J. W. FOSTER. 1996. Biosynthesis and recycling of NAD. Pp. 721-730 *in* F. C. Neidhardt, ed. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. ASM Press, Washington, D.C.

PHILLIPS, P. C. 2005. Testing hypotheses regarding the genetics of adaptation. Genetica 123:15-24.

PONCE, E., N. FLORES, A. MARTINEZ, et al. 1995. Cloning of the two pyruvate kinase isoenzyme structural genes from Escherichia coli: the relative roles of these enzymes in pyruvate biosynthesis. J. Bacteriol. 177:5719-5722.

POSTMA, P. W., J. W. LENGELER, and G. R. JACOBSON. 1996. Phosphoenolpyruvate:carbohydrate phosphotransferase systems. Pp. 1149-1173 *in* F. C. Neidhardt, ed. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. ASM Press, Washington, D.C.

PRICE, T., M. TURELLI, and M. SLATKIN. 1993. Peak shifts produced by correlated response to selection. Evolution 47:280-290.

PURUGGANAN, M. D. 2000. The molecular population genetics of regulatory genes. Mol. Ecol. 9:1451-1461.

RAUSHER, M. D. 1988. Is coevolution dead? Ecology 69:898-901.

REMINGTON, D. L., and M. D. PURUGGANAN. 2003. Candidate genes, quantitative trait loci, and functional trait evolution in plants. Int J Plant Sci 164:S7-S20.

REMOLD, S. K., and R. E. LENSKI. 2001. Contribution of individual random mutations to genotype-by-environment interactions in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 98:11388-11393.

RICE, W. R. 1989. Analyzing tables of statistical tests. Evolution 43:223-225.

ROSE, M. R. 1991. Evolutionary biology of aging. Oxford University Press, Oxford.

ROZEN, D. E., J. A. G. M. DE VISSER, and P. J. GERRISH. 2002. Fitness effects of fixed beneficial mutations in microbial populations. Curr. Biol. 12:1040-1045.

SANJUAN, R., A. MOYA, and S. F. ELENA. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc. Natl. Acad. Sci. USA 101:8396-8401.

SARUBBI, E., K. E. RUDD, and M. CASHEL. 1988. Basal level adjustment shown by new *spoT* mutants affect steady state growth rates in *rrnA* ribosomal promoter regulation in *Escherichia coli*. Molelcular and General Genetics 213:214-222.

SAS INSTITUTE. 1999. SAS Version 8, Cary, N.C.

SCHEFFÉ, H. 1959. The analysis of variance. Wiley, New York.

SCHEINER, S. M. 1993. Genetics and evolution of phenotypic plasticity. Annu. Rev. Ecol. Syst. 24:35-68.

SCHLUTER, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1888-1895.

---. 2000. The Ecology of Adaptive Radiation. Oxford University Press, Oxford.

SCHLUTER, D., E. A. CLIFFORD, M. NEMETHY, et al. 2004. Parallel evolution and inheritance of quantitative traits. Am. Nat. 163:809-822.

SCHNEIDER, D., E. DUPERCHY, E. COURSANGE, et al. 2000. Long-term experimental evolution in Escherichia coli. IX. Characterization of insertion sequence-mediated mutations and rearrangements. Genetics 156:477-488.

SERVICE, P. M., and M. R. ROSE. 1985. Genetic covariation among life-history components: the effect of novel environments. Evolution 39:943-945.

SHIGENOBU, S., H. WATANABE, M. HATTORI, et al. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407:81-86.

SIMPSON, G. G. 1953. The major features of evolution. Columbia University Press, New York.

SOKAL, R. R., and F. J. ROHLF. 1995. Biometry. Freeman, New York.

STEPPAN, S. J., P. C. PHILLIPS, and D. HOULE. 2002. Comparative quantitative genetics: evolution of the G matrix. TREE 17:320-327.

SYSTAT. 2002. SYSTAT Version 10.2, Richmond, CA.

TRAVISANO, M. 1997. Long-term experimental evolution in *Escherichia coli*. V. Environmental constraints on adaptation and divergence. Genetics 146:471-479.

TRAVISANO, M., and R. E. LENSKI. 1996. Long-term experimental evolution in *Escherichia coli*. IV. Targets of selection and the specificity of adaptation. Genetics 143:15-26.

TRAVISANO, M., F. VASI, and R. E. LENSKI. 1995. Long-term experimental evolution in *Escherichia coli*. III. Variation among replicate populations in correlated responses to novel environments. Evolution 49:189-200.

VASI, F., M. TRAVISANO, and R. E. LENSKI. 1994. Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. Am. Nat. 144:432-456.

VASI, F. K., and R. E. LENSKI. 1999. Ecological strategies and fitness tradeoffs in Escherichia coli mutants adapted to prolonged starvation. J. Genet. 78:43-49.

VIA, S., and R. LANDE. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. Evolution 39:505-522.

WEBER, K. E. 1996. Large genetic change at small fitness cost in large populations of Drosophila melanogaster selected for wind tunnel flight: Rethinking fitness surfaces. Genetics 144:205-213.

WERNEGREEN, J. J., A. B. LAZARUS, and P. H. DEGNAN. 2002. Small genome of Candidatus *Blochmannia*, the bacterial endosymbiont of *Camponotus*, implies irreversible specialization to an intracellular lifestyle. Microbioology 148:2551-2556.

WHITLOCK, M. C. 1996. The Red Queen beats the Jack-of-All-Trades: the limitations on the evolution of phenotypic plasticity and niche breadth. Am. Nat. 148:S65-S77.

WHITLOCK, M. C., P. C. PHILLIPS, F. B.-G. MOORE, et al. 1995. Multiple fitness peaks and epistasis. Annu. Rev. Ecol. Syst. 26:601-629.

WILKE, C. O. 2004. The speed of adaptation in large asexual populations. Genetics 167:2045-2053.

WILKE, C. O., and C. ADAMI. 2002. The biology of digital organisms. TREE 17:528-532.

WINGREEN, N. S., J. MILLER, and E. C. COX. 2003. Scaling of mutational effects in models for pleiotropy. Genetics 164:1221-1228.

WRIGHT, S. 1968. Evolution and the Genetics of Populations. University of Chicago Press, Chicago.

---. 1977. Evolution and the Genetics of Populations. University of Chicago Press, Chicago.

---. 1984. Evolution and the Genetics of Populations. University of Chicago Press, Chicago.

ZIJLSTRA, W. G., M. J. STEIGENGA, P. M. BRAKEFIELD, et al. 2003. Simultaneous selection on two fitness-related traits in the butterfly *Bicyclus anynana*. Evolution 57:1852-1862.