8

Phenotypic and Genomic Evolution during a 20,000-Generation Experiment with the Bacterium *Escherichia coli*

*Richard E. Lenski**
Center for Microbial Ecology
Michigan State University
East Lansing, Michigan 48824

*Acknowledgments: Many individuals have contributed to this long-term experiment. Rather than trying to list everyone here, I will let the literature cited record many of the contributions of the graduate students, postdoctoral associates, and colleagues who have worked on this project. However, I want to emphasize the contributions of a few individuals in particular. Michael Rose played an important role in getting this project started, as his own long-term study selecting for delayed senescence in fruitflies provided an elegant model to emulate. The *E. coli* experiment has depended on the dedication of three outstanding technicians over the years: Sue Simpson, Lynette Ekunwe, and Neerja Hajela. And Madeleine Lenski has tolerated my long-term "affair" with this experiment. This research has been supported by the National Science Foundation (currently grant DEB-9981397) and by Michigan State University. James Crow and Jules Janick provided many helpful comments on this paper.

Plant Breeding Reviews, Volume 24, Part 2, Edited by Jules Janick
ISBN 0-471-46892-4

I. INTRODUCTION

This paper reviews another long-term selection experiment, one that is both shorter and longer than the 100 years of selection for oil and protein content in maize that is the main focus of this volume. In 1988, 12 populations of the bacterium *Escherichia coli* were founded from the same strain, and these have been propagated in a simple, defined laboratory environment ever since. Each day, the populations are diluted in fresh medium, undergo about 6.6 generations of binary fission before they exhaust the limiting resource, and then must wait until their "springtime" appears again the next day.

This review summarizes some interesting changes and dynamics, both phenotypic and genomic, that occurred in these populations through generation 20,000. For an annual plant 20,000 generations would of course require some 20,000 years; for humans 20,000 generations would span 400,000 years, assuming an average generation of 20 years. This ability to observe evolution in action over many generations is an obvious benefit of studying bacteria. In fact, the experiment recently passed 30,000 generations, but the bacteria evolve faster than we can study them, and generation 20,000 represents the last milestone at which we systematically studied the populations. Rather than taking the same measurements at intervals, we have repeatedly pushed our analyses in new directions. Thus, some changes were analyzed through 2,000 generations, others through 10,000 generations, and still other changes through 20,000 generations. Another wonderful feature of bacteria for evolutionary research is that we can store entire populations frozen, and resurrect them whenever we wish, such that we can gather more data about earlier generations if we later find other aspects of the evolving bacteria that we want to measure and study.

A. Motivating Questions

This long-term evolution experiment with *E. coli* has several goals. First, the experiment aims to measure the dynamics of evolutionary change. Is evolution strictly gradual, or are there episodes of more rapid change even in a constant environment? Does the initial rate of change continue indefinitely, or is some limit eventually reached? And how might the dynamic patterns depend on which particular phenotypic or genetic traits are measured?

Second, this experiment seeks to examine the repeatability of evolution by having 12 replicate populations, all of them founded from the same ancestor and maintained in the same environment. Which phenotypic and genomic aspects of evolutionary change are repeatable in this system, and which are haphazard? How can we understand the causes of parallelism and divergence of replicate lines? The issue of the predictability of evolutionary change—or lack thereof—has long been of interest. The question was well captured by the late paleontologist Steven Jay Gould (1989) in a thought experiment: "*I call this experiment 'replaying life's tape.' You press the rewind button and, making sure you thoroughly erase everything that actually happened, go back to any time and place in the past—say to the seas of the Burgess Shale. Then let the tape run again and see if the repetition looks at all like the original . . .*" Gould went on to say, however, that "*The bad news is that we can't possibly perform the experiment.*" Of course, we could never run an experiment on the vast temporal and spatial scales imagined by Gould. But on much smaller scales, our experiment with *E. coli* allows us to address the same question. We do so by allowing 12 scenarios to play out simultaneously, rather than sequentially as the notion of replay implies, but the issue of repeatability is fundamentally the same.

Third, the long-term evolution experiment offers the opportunity to integrate data on phenotypic and genetic changes. Are the dynamics of phenotypic and genomic change concordant? If there are discrepancies, why do they occur? If certain phenotypic traits evolve in parallel across the replicate populations, does this imply parallelism at the level of mutations, genes, or pathways? Have the bacteria become ecological specialists as they adapted to the monotonous selective regime? If so, what are the roles of pleiotropic tradeoffs versus accumulation of neutral mutations in causing specialization? How much phenotypic and genomic evolution can occur in 20,000 generations, and how well do the observed rates accord with rates inferred from comparative data? It may seem that linking genomic and phenotypic changes should be quite easy in a model system like *E. coli*, with its relatively small (and now

totally sequenced) genome, unicellular structure, and decades of service as a model system for molecular genetics. But linking phenotypic and genomic changes remains a real challenge for several reasons: (1) we are looking for a relatively few mutations in a genome that is several million base-pairs long; (2) most phenotypic effects of interest in this experiment are subtle quantitative changes, as opposed to the knock-outs causing losses of function usually studied by geneticists; (3) the absence of genetic markers and Mendelian recombination in the base population (which was clonal) force us to perform precise genetic manipulations to establish the effects of particular mutations on phenotypes of interest; and (4) the functions of many genes remain unknown and, even for genes that have been well studied, the extent of interactions with other genes is poorly known. Despite these difficulties, we have made substantial progress in linking genomic and phenotypic changes in the long-term *E. coli* lines. Some of this work has been published in the primary literature, but much more is still in progress. To avoid precluding later publication of this on-going work, the present review will present only those findings that have been published while describing on-going analyses in general terms.

An important aspect that impinges on all these questions is that our experimental system differs profoundly from the more familiar plant and animal models used in most other selection experiments. Bacteria are haploid, and they reproduce asexually. Many bacteria in nature have parasexual processes that allow varying degrees of recombination, but the *E. coli* strain used in this experiment lacks the potential to engage in these processes. Hence, these populations are strictly asexual. [In a separate experiment, we examined some of the consequences of allowing one of these parasexual processes to occur in association with introducing genes from another strain of *E. coli* (Souza et al. 1997).] Moreover, each replicate line in the long-term experiment was founded from a single clone (in fact, one cell) and hence there was no standing variation at the start of the experiment. New mutations thus provide the sole source of genetic variation available to selection. Therefore, any parallelism that we observe necessarily involves both the independent origin and fate of variants, as opposed to mere sorting of preexisting variants that are identical by descent. This dependence on new variation is one reason I prefer to call this an evolution experiment rather than a selection experiment. Also, selection in this long-term *E. coli* experiment is natural selection, as opposed to artificial selection as practiced by plant and animal breeders. By that I mean the environment selects; we only provide the environment (in this case, a simple and artificial one), but we do not choose particular individual cells to reproduce on the basis of their

phenotypes. Finally, I would mention that bacteria are of tremendous importance, not only as pathogens but as fundamental players in every natural and managed ecosystem. Yet, until recently bacteria have been largely ignored by evolutionary biologists, who perhaps regarded them as tools of molecular biology and lacking in the obviously colorful phenotypes of many plants and animals. The importance of bacteria in nature, along with their very different genetic systems, combine to make them an interesting area for evolutionary research.

B. Overview of the Experimental System

This section describes some important features of the experimental system. I have tried to limit the microbiological jargon and tedious details, which can be found in the various primary papers cited in this review. Instead, I want to give the reader a feel for how the experiment is performed and how the resulting evolutionary changes are discerned. Let me begin by summarizing several features that make bacteria in general, and *E. coli* in particular, powerful for studying evolution by an experimental approach. (I have chosen to pursue an experimental approach to evolution for two reasons: I enjoy experiments, and I feel that evolutionary biologists tend to underutilize the experimental approach. Of course, I realize that evolutionary biology is, first and foremost, an historical science and must rest primarily on comparative and paleontological data.)

1. Experimental Advantages of *E. coli*. First, *E. coli* are very easy to grow and count. Second, they grow in simple environments that are easy to control and manipulate, for example by varying culture temperature or resource supply. Third, *E. coli* have rapid generations, allowing experiments to last hundreds or thousands of generations; and they can have large populations, providing a substantial input of mutational variation. The resulting supply of mutations is quantified in Section I.B.4 below.

Fourth, one can preserve cells in an ultra-low freezer, and later resurrect them as needed. Thus, one can compare ancestral and derived forms directly, at any time, without relying on the possibility that protocols or conditions have subtly changed over the years. Not only can one store the original ancestor, but one can store samples from intermediate times in the experiment. These samples contain not mere clones but, in effect, the entire population (less the fraction used to propagate the population). Therefore, if one discovers that something interesting has changed in some later generation, then one can go into the "frozen

fossil record" to discover when the difference emerged, what allele frequencies were at various times, and so on. And when an accident disrupts the experiment (say, a malfunctioning incubator or contamination), one can simply restart the populations from the most recent samples. (When I switched from studying insects to bacteria, I was obviously attracted to the rapidity of bacterial generations. But I realize now that the ability to store and revive organisms and populations is just as important for this type of research.)

Fifth, the *E. coli* strain used in this experiment is strictly clonal. As mentioned earlier, many bacteria undergo parasexual processes, but the strain used in this study lacks the requisite mechanisms to do so. As a consequence of this clonality, one can place a genetic marker in a particular background and have it remain there. By placing different states of an immobile marker in the replicate populations, we protect against the possibility that inadvertent cross-contamination could go undetected and thus not be corrected by re-starting lines. Also, the immobile marker allows us to mix two populations, which brings us to the next advantage. Sixth, one can directly measure the mean fitness of derived bacteria relative to their ancestor. Fitness assays are head-to-head competition experiments, typically between a derived population and its ancestor, except the ancestor carries a neutral genetic marker that allows it to be readily distinguished on an appropriate medium. Section I.B.3 provides more detailed information on how we define and measure relative fitness.

Seventh, *E. coli* has long served as a model organism in genetics, genomics, molecular biology, biochemistry, and cell physiology. Hence, there is a wealth of information about this organism that can potentially help us interpret our own findings (Neidhardt et al. 1996; Blattner et al. 1997).

2. Experimental Design and Conditions. The experiment includes 12 replicate populations, all founded from the same ancestor, except for a single genetic marker described below. To start the experiment, the bacteria were plated as single colonies, each derived from a single cell, and then a separate colony was used to inoculate each population. Thus, each population was founded from a single cell, and hence each population has depended entirely on new mutations for its subsequent evolution (Lenski et al. 1991).

Six of the populations are unable to catabolize the sugar arabinose. The other six have a point mutation in the *ara* operon that allows them to grow on arabinose. The Ara marker is selectively neutral under the conditions of our experiment (as shown by many control fitness assays),

but it serves two purposes. First, as described in Section I.B.3, it allows us to distinguish two genotypes or populations when they are deliberately mixed to measure their relative fitness. Second, in the course of propagating our lines, we always alternate between Ara^- and Ara^+ lines. Therefore, any inadvertent cross-contamination event would introduce the wrong marker state, which can then be detected and corrected by restarting the affected populations from frozen samples. This protection has allowed us to be confident that even parallel changes are in fact evolutionarily independent (although it has become less critical as our genetic analyses have now identified other molecular markers that evolved during the course of the experiment and uniquely identify each population).

The bacteria live in a liquid, buffered, minimal-salts medium supplemented with glucose as the sole source of carbon and energy (Lenski et al. 1991). Glucose is limiting to bacterial density, and it is supplied at a concentration of 25 μg/ml which, although much lower than that used in most microbiological experiments, allows the bacteria to reach a density of about 5×10^7 cells per ml when the glucose is depleted. Each culture is 10 ml, so that the final population size is about 5×10^8 cells. The cultures are held in small Erlenmeyer flasks and incubated at 37°C.

Every day, each population is serially transferred by diluting 0.1 ml into 9.9 ml of fresh medium. This basic rhythm has continued, day in and day out, for over a decade now (with a few interruptions as described below). The 100-fold dilution and re-growth allow $\log_2 100 = 6.64$ generations of binary fission per day. In fact, the bacteria grow and deplete the available glucose in the first eight hours or so, and then spend the remainder of the day in stationary phase. (In principle, we could transfer the bacteria more often, or dilute them more than 100-fold each day, and thereby increase the number of generations. However, daily transfers are logistically simpler, and imposing a larger dilution would reduce the effective population size.) To the extent that there is cell death after the glucose is depleted, the number of generations might be even more than 6.64 per day; in fact, however, the founding strain experiences no appreciable mortality over the course of a day of starvation (Vasi et al. 1994). Fifteen days, therefore, allow about 100 generations, and some 2,400 generations occur in a year.

Samples of all the evolving populations have been stored at intervals, initially every 100 generations and later at 500-generation intervals. Samples are stored at –80°C, with glycerol added as a cryoprotectant. We have always been readily able to revivify bacteria, even after more than a decade in storage. There is no evidence from either performance measures or genomic analyses that the bacteria have mutated during their

storage. An important feature of our storage regimen is that we freeze entire populations, and not just individual clones (although we sometimes store clones as well). That is, after a population has been serially transferred, most of the population is left behind; we then add glycerol to that almost complete population (less the 1% that was transferred) and store it away. Thus, we can recover the entire population, not just individual clones, for later analysis or to restart the population if needed.

Over the course of many years, accidents and disruptions of various sorts can and do happen. For example, the experiment was halted for some months when I moved to Michigan State University from the University of California, Irvine, where it began. The experiment has also been restarted from the frozen samples on several occasions when contamination occurred (including cross-contamination of one line by another). Despite disruptions, the bacteria have undergone more than 30,000 generations since the experiment began on 15 February 1988. We are limited more by the time and effort required to analyze the bacteria in meaningful and new ways than by the number of bacterial generations elapsed.

Prior to making any measurements, all clones or populations that will be studied are stored in the freezer, removed simultaneously, and then acclimated to the culture conditions under which the assays will be performed. In effect, all measurements are therefore performed in a "common garden" following several generations of acclimation that eliminate any non-heritable effects of prior growth conditions. In this way, we can be confident that significant differences we observe between clones or populations, whether in fitness or some other property, reflect underlying genetic differences (even before we have identified the responsible mutations).

I am often asked about the ancestral strain and what environment it had evolved in prior to the start of the long-term experiment. The so-called B strain used in the long-term evolution experiment had already been used in laboratories for several decades, during which time it was grown in many different media and stored for periods, either under starvation conditions at room temperature (the old-fashioned method) or in a freezer where cells are metabolically inactive (the modern approach). Some adaptive evolution undoubtedly occurred in the laboratory prior to the start of the long-term experiment, although the opportunity for adaptation to these conditions was certainly small compared with the millions of years that its *E. coli* ancestors spent in, and moving between, the digestive tracts of their vertebrate hosts. More importantly for our purpose here, the environment in the long-term experiment is

clearly novel from the perspective of the bacteria, not in the sense that any one aspect (e.g., presence of glucose) has been encountered for the first time, but rather in the combination of all the factors, including their uniformity from day to day.

3. Defining and Measuring Fitness. We measure relative fitness by competing two clones or populations against one another. These competitions are performed as separate experimental assays, and do not impinge on the continuing propagation of the long-term lines themselves. By counting the number of the two competing types at the beginning and the end of a competition experiment, we can calculate the net growth rate that the bacteria of each type achieved while they competed with the other type for the common pool of resources (Lenski et al. 1991). In our standard serial transfer regime, differences in net growth rate could reflect, in principle, not only differences in exponential growth but also differences in the duration of the lag prior to the commencement of growth, the effect of diminishing resources on growth, and survival after the resources have been depleted. Relative fitness is then simply defined as the ratio of the two competitors' net growth rates.

Using hypothetical numbers for simplicity, consider the following example. An evolved population and genetically marked ancestor (see below) are mixed at a 1:1 ratio, with the initial density of each 2×10^5 cells per ml. Over the course of a day, they collectively grow 100-fold because the assay is performed using the same glucose-limited medium and dilution factor as in the long-term evolution experiment. However, the ratio of the two competitors changes during the competition, such that at the end of 24 h the ratio is 3:1, with final densities of the evolved and ancestral types being 3×10^7 and 1×10^7 cells per ml, respectively. The realized net growth rate of the evolved competitor is calculated as $\log_e [(3 \times 10^7)/(2 \times 10^5)] \cong 5.01$ per day, while the corresponding rate for the ancestor is $\log_e [(1 \times 10^7)/(2 \times 10^5)] \cong 3.91$ per day. Thus, the fitness of the evolved population relative to the ancestor is $5.01/3.91 \cong 1.28$, which is a dimensionless quantity because the units cancel. Notice that this quantity is smaller than the 3-fold change in the relative abundance of the two competitors from start to finish, because the competition assay compounds the difference in growth rates over several generations.

In this review, I will often discuss the mean fitness of an evolved population relative to the ancestor, usually as measured under the same environmental conditions as used in the long-term evolution experiment (unless otherwise stated). However, one can also compete an evolved population against the ancestor under some other conditions than those

used in the evolution experiment, and thereby examine how adaptation to one environment produced correlated changes in fitness measured in a different environment.

A key operational issue is how we distinguish our two competitors. For example, when we mix and compete an evolved population against the ancestor, how can we tell them apart in order to calculate their net growth rates and hence fitness? The Ara genetic marker that was mentioned earlier provides the key. Ara^- and Ara^+ cells produce red and white colonies, respectively, on an appropriate agar medium. Thus, we can tell any two competitors apart provided that one is Ara^- and the other Ara^+. For example, when measuring the fitness of an evolved population relative to the ancestor, we always compete the evolved population against the ancestral clone that carries the opposite marker. The six Ara^- populations compete against the Ara^+ ancestor, and the six Ara^+ populations compete against the Ara^- ancestor. The Ara marker has been shown repeatedly to be selectively neutral in our standard glucose-limited environment, and whenever we run experiments in other environments suitable controls are included. Finally, although we distinguish the two competitors on an agar-based medium, that medium is used only to distinguish and count the bacteria; the actual competition occurs in the liquid environment. (In principle, two competitors might differ in their plating efficiency, that is the fraction of cells in liquid that produce colonies on the agar medium. In fact, however, any such difference has no effect on calculated growth rates, and hence relative fitness, provided only that each competitor's plating efficiency is the same in the initial and final samples. To ensure that this condition is fulfilled, we estimate both initial and final densities at the same fixed point in the population cycle, when all cells are in stationary phase.)

4. Number of Mutations per Population. Just how many mutations have occurred in any one of the evolving populations over the course of 20,000 generations? To answer this question, it is critical to distinguish between the number of mutational events, the vast majority of which are lost from the population, and the number of mutations substituted in the population by the combined effects of natural selection and random drift.

Taking each definition in turn, the number of mutational events across the entire genome is equal to the product of the population size, number of generations, base-pair mutation rate, and genome size. The size of each population fluctuates daily between about 5×10^6 and 5×10^8 owing to the dilution and subsequent re-growth. The effective popula-

tion size, expressed in terms of the process of origin and substitution of beneficial mutations, is approximately equal to the minimum (bottleneck) population size times the number of generations between minimum and maximum sizes: $5 \times 10^6 \times \log_2 100 \approx 3 \times 10^7$ (Lenski et al. 1991). The mutation rate for *E. coli*, assuming normal DNA repair and editing (which are functional in the ancestral strain), has been estimated to be about 5×10^{-10} per base-pair (Drake et al. 1998), although evidence from our own experiment suggests a somewhat lower ancestral rate of about 1×10^{-10} per base-pair (Lenski et al. 2003). The genome length for *E. coli* is about 5×10^6 base-pairs (Blattner et al. 1997). Thus, the total number of mutations per population during 20,000 generations is expected to be approximately 1.5×10^9 using the higher mutation rate, and about 3×10^8 using the lower rate. That is, several hundreds of millions of point mutations have appeared in each population, even after adjusting for the transfer bottleneck. In addition to these point mutations, other types of spontaneous mutation can also occur, including insertions, deletions, and inversions; many of these larger events involve insertion sequences, or IS elements, that are present in most bacteria.

That is certainly a lot of mutations. In fact, with a genome length of 5×10^6 base-pairs and three alternative base-pairs per position, only 1.5×10^7 point mutations are even possible. Thus, each population has had most point mutations represented many times over. Of course, this calculation does not imply that every possible sequence has ever existed—far, far from it—because each mutation occurs against a relatively few genetic backgrounds. The backgrounds change as the population evolves, but as we shall see below, that vast majority of the genome does not change in the majority of cells.

At first glance, this mutational redundancy might suggest that each population should have "tried" and substituted any mutation that is beneficial generally across all backgrounds. But while the populations should repeatedly try all beneficial mutations, that does not ensure their substitution. As Haldane (1927) showed, most beneficial mutations are lost by random drift before they become common enough for selection to drive them to fixation; the probability of substitution of a beneficial mutation is only about $2s$, where s is its selective advantage. Thus, a mutation that confers a 10% advantage will be lost by drift about 80% of the time, whereas one with only a 0.5% advantage will disappear 99% of the time. Twenty-fold mutational redundancy should overcome drift loss in the former case, but not in the latter. The situation is made even more complicated, and the fixation probabilities further reduced, by the fact that the populations studied here are asexual, which leads to a

phenomenon called clonal interference (Muller 1932; Gerrish and Lenski 1998). In essence, clonal interference occurs because beneficial mutations may arise in two or more different clones but, in the absence of recombination, only one of them can ultimately go to fixation. In general, mutations with only weak beneficial effects, even those lucky enough to escape drift loss, require a very long time to become fixed or even numerically dominant, and thus a clone bearing a weakly beneficial mutation will usually be out-competed by some more beneficial mutation that appears on another background. Thus, the $2s$ probability of fixation calculated by Haldane is, in fact, an upper limit.

Given that a billion or so mutations appear in each population during 20,000 generations, how many are likely to have been substituted? It is difficult to give a precise answer, but we can consider two special cases in order to give us a rough sense of the number of fixations that could reasonably occur by selection and random drift in this amount of time. Let us consider selection first. We will assume that the maximum selective benefit of any mutation is 10%, which accords well with experimental findings that will be given later. Given that a mutation first appears in a single individual, and considering the case where it is lucky enough to survive drift loss or clonal interference, it will take about 250 generations for the mutant genotype to achieve a frequency of 50% in the population, at which point it is more likely to get the next beneficial mutation that comes along instead of competing with it (Lenski et al. 1991). Assuming that beneficial mutations are neither so rare as to require a very long waiting period given the population size, nor so common that beneficial mutations often occur in minority subpopulations, then we can imagine "stringing together" 20,000/250 = 80 beneficial substitutions, each of 10% effect, in a 20,000-generation period. In fact, however, the real number of beneficial substitutions is almost certainly much less than this number, as evidenced by waiting periods between substitutions and a pronounced deceleration in the rate of adaptation over the course of the experiment. I would hazard the estimate that between 10 and 20 beneficial mutations have been substituted in each population.

Turning to drift, the steady-state rate of substitution of neutral mutations is equal simply to their genome-wide rate of occurrence (Kimura 1983). Using the genome size and even the higher base-pair mutation rate given earlier, and assuming that most mutations are neutral, one expects a neutral substitution rate of only about 0.0025 per generation. Over 20,000 generations, that rate would allow about 50 neutral substitutions per evolving population. The number would be somewhat lower if a substantial fraction of mutations are deleterious, but that does not appear

to be the case (Kibota and Lynch 1996). The situation is complicated (but less so than it first seems) by the fact that a drift substitution requires on the order of N generations, where N is population size. Given that the bacterial population size is in the millions, and the experiment has run for only thousands of generations, there has not been enough time for a drift substitution starting from a homogenous initial state. But on-going selection of beneficial mutations actually simplifies things, because each selective substitution traces back to an individual cell in an asexual population, and hence the drift-effective size of these populations is much smaller than their nominal size. (This drift-effective population size is a different quantity from the effective population size that matters for the origin and fate of beneficial mutations themselves.) These selective substitutions participate in driving out neutral mutations that occur in other clones, but the occasional neutral mutation will hitchhike to fixation with the beneficial mutation that appears in the same genome. By lowering the drift-effective population size, this selective process dramatically shortens the time required for a particular neutral mutation to be substituted. However, this same selection does not affect the overall rate of neutral substitutions, which is independent of population size (Kimura 1983).

Combining beneficial and neutral substitutions, it seems probable that fewer than 100 mutations could have been substituted in each population, out of the billion or so that occurred. Thus, the bacterial genomes will have changed very little in the broad scheme of things, which is reassuring given that a decade is a mere "drop in the bucket" in terms of molecular evolution. It is also why the hunt for mutations is difficult in the context of evolving populations that started with no standing variation. Finally, let me mention in advance that several, but not all, of the populations evolved genetic defects in their DNA repair pathways that caused a large increase in their mutation rate. Therefore, the number of mutations, especially neutral ones, substituted in those populations should be greater than suggested by the calculations above. We will return to this interesting complication in Section II.C.

II. PHENOTYPIC AND GENOMIC EVOLUTION

In principle, it might be nice to separate sections on phenotypic and genomic evolution, but such separation is not feasible or appropriate for two reasons. First, the distinction is imprecise because certain mutations are discovered by virtue of having observed particular phenotypic changes. Second, one of the main goals of the genomic analyses is to

examine the coupling between phenotypic and genomic changes, in terms of both overall dynamics and the effects of particular mutations on phenotypic changes. Nonetheless, the overall trend of this main section will be from the phenotypic to genomic levels.

A. Relative Fitness

In a number of experiments over the years, we have measured the temporal trajectory for fitness of the evolving populations relative to their common ancestor. Several important features of the data are as follows. First, as shown in Fig. 8.1, the average fitness gain for the 12 populations after 20,000 generations is about 70% (Cooper and Lenski 2000). That is, during competition between an evolved population and its ancestor, the evolved bacteria undergo approximately 1.7 cell doublings for every cell doubling by the ancestor.

Second, the rate of improvement in fitness has tended to decelerate over time. Lenski et al. (1991) found that the rate of improvement was lower between generations 1,000 and 2,000 than it was between 0 and 1,000 generations in all 12 populations. The average gain in the earlier period was about twice that in the later period. Lenski and Travisano

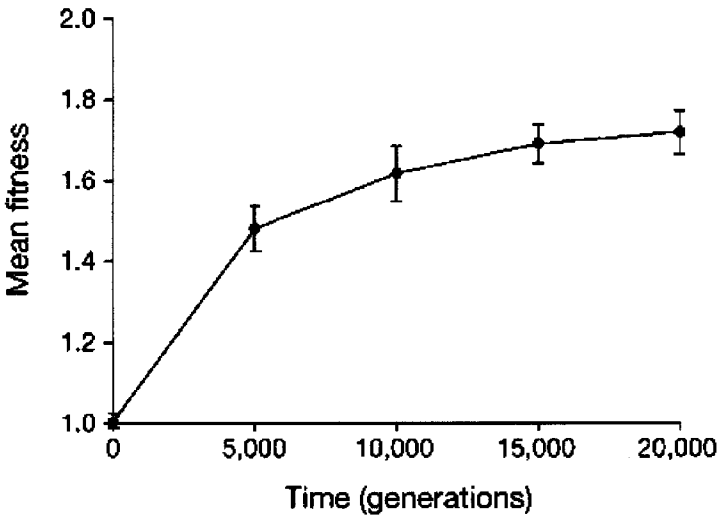

Fig. 8.1. Trajectory of mean fitness, averaged over 12 evolving populations of *E. coli*, through 20,000 generations. Fitness is expressed relative to the ancestor, and it measures the ratio of growth rates realized during direct competition in the selective environment. Error bars are 95% confidence intervals. Source: Cooper and Lenski 2000. Reprinted with permission of *Nature*.

(1994) extended this finding of decelerating adaptation to 10,000 generations. Most recently, Cooper and Lenski (2000) reported fitness trajectories through 20,000 generations, and calculated that the average rate of improvement relative to the ancestor between 15,000 and 20,000 generations was only about one-tenth the average rate during the first 5,000 generations. Nonetheless, there was significant improvement even during this last interval, indicating that adaptation had slowed down but not stopped.

Third, the fitness trajectory for individual populations is not a smooth curve, but instead follows a step-function (Lenski et al. 1991; Lenski and Travisano 1994), as illustrated in Fig. 8.2. The step-like increases are produced when beneficial mutations sweep through an evolving population, and the observed dynamics accord well with mathematical models of the process of genetic adaptation in asexual populations (Lenski et al. 1991; Gerrish and Lenski 1998; see also Orr 1998). [The step-like trajectory also shows that punctuated evolution can occur very simply, although it has been debated whether these data are relevant to the punctuated-equilibrium model based on the fossil record (Gould and

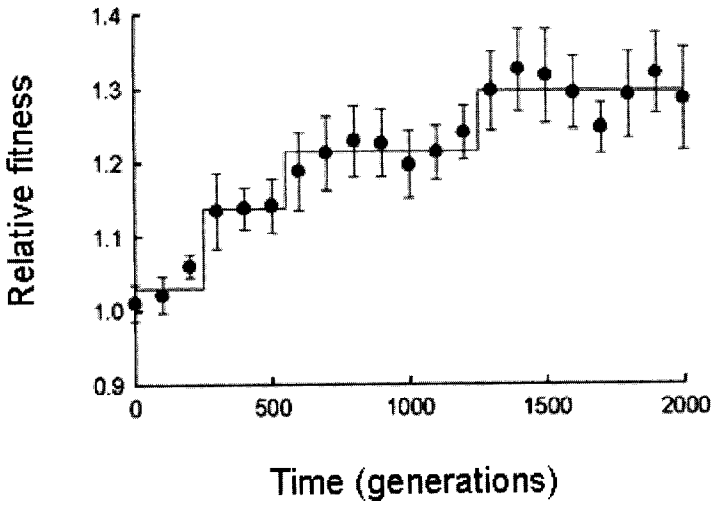

Fig. 8.2. Trajectory of mean fitness in one evolving population through 2,000 generations. The step-like function provides a better fit than simpler models. Each step presumably represents the spread of a beneficial mutation of large effect. The overall trajectory appears step-like owing to the underlying dynamics of mutation and selection in asexual populations, and given the 100-generation sampling interval. Error bars are 95% confidence intervals. Source: Lenski and Travisano 1994. Copyright 1994 National Academy of Sciences, U.S.A.

Eldredge 1977; Elena et al. 1996; Coyne and Charlesworth 1996; Gould 2002).] The steps are observed only in the trajectories for individual populations, because beneficial mutations occur stochastically and replicate populations are not synchronized. Also, resolving individual steps requires intensive replication of the fitness measurements at a relatively fine temporal scale, and the larger steps that occur early in the evolution experiment are therefore easiest to discern. The early steps produce fitness gains of roughly 10% relative to the ancestor. Given the population size in the experiment, a mutation that appears in an early generation and confers a 10% advantage (and escapes loss by genetic drift) will still be a small minority after 200 generations, but by 300 generations it will comprise the vast majority. Thus, if fitness assays are made every 100 generations, the trajectory will appear to jump between 200 and 300 generations. Longer periods for ascendancy are required if there is a substantial waiting time for a beneficial mutation to appear, and for mutations that confer smaller benefits. Also, owing to the asexuality of the evolving populations, each beneficial mutation that is substituted effectively purges the standing genetic variation in the population, as noted before in the context of neutral mutations. This purging includes other beneficial mutations that appeared in other backgrounds, but which have smaller beneficial effects than the one that ultimately prevails, a phenomenon called clonal interference (Muller 1932; Gerrish and Lenski 1998).

Fourth, the fitness trajectories for the 12 replicate populations are very similar in their overall form and the extent of improvement, but they are not identical. Lenski et al. (1991) showed that, during the first 2,000 generations, the among-population variance component for fitness was significant and corresponded to a standard deviation of about 3%, or approximately one-tenth the change in mean fitness over that period. Lenski and Travisano (1994) also found fitness variation of several percent among the replicate populations through 10,000 generations. Although Cooper and Lenski (2000) did not specifically address this issue, an analysis of their data indicates that significant among-population variation in mean fitness persisted throughout 20,000 generations. Thus, there exists subtle variation in performance among the 12 replicate populations, despite their similar trajectories.

Fifth, in addition to the variation in mean fitness among the replicate populations, there also exists significant variation among clones sampled from the same population. Lenski et al. (1991) quantified the within-population variance for fitness in the first 2,000 generations. They found that the observed variance was not significantly different from the variance that must be produced by the on-going substitution of

beneficial mutations, a quantity that they calculated using Fisher's fundamental theorem. Elena and Lenski (1997) performed a similar analysis at 10,000 generations. They found that the within-population variation in fitness had not changed much since 2,000 generations. However, because the rate of adaptation had decelerated so much over that time, the on-going substitution of beneficial mutations no longer provided a sufficient explanation for the observed level of variation. They considered two alternative explanations, according to which the within-population variation in fitness was maintained by the balance between deleterious mutations and selection, or by frequency-dependent selection. The former also was insufficient to account for the observed variation. However, they observed a tendency for clones to have advantages when rare relative to other clones in their population, which could account for this variation. In five of the six populations examined, the average advantage-when-rare tended to be very small, about 1–2%. However, in one population the average clonal advantage-when-rare was near 7%. Rozen and Lenski (2000) studied this population in detail and found that it contained two distinct ecotypes that stably coexisted. One of the two ecotypes was superior at exploiting the glucose that was provided in the medium, but it also secreted a metabolite that the other ecotype could better use. Thus, frequency-dependent selection enters into the dynamics even in this very simple system. However, it should also be emphasized that the magnitude of these frequency-dependent effects are usually quite small compared with the large improvements relative to the ancestor.

Sixth, and finally, one can ask whether the pronounced deceleration in the rate of fitness improvement is an artifact of performing competitions between the evolved lines and their ancestor. For example, in a study of evolving yeast populations, Paquin and Adams (1982) uncovered non-transitive competitive interactions. In their study, the fitness of evolved yeast clones measured relative to the original ancestor sometimes declined, whereas a clone's fitness measured relative to its own immediate predecessor always increased. Various indirect lines of evidence argue against the importance of this phenomenon in the long-term experiment with *E. coli*, including the relatively weak and negative frequency-dependent effects described above, as well as the observation that absolute fitness components, including exponential growth rate and the duration of the lag prior to growth, consistently improved during the evolution experiment (Vasi et al. 1994). We also performed some 300 competitions between samples from various generations of the same population, to test the possible contribution of non-transitive effects to the fitness trajectory (De Visser and Lenski 2002). In short, these data

accord very well with transitive interactions. Therefore, the decelerating trajectory of mean fitness measured relative to the common ancestor indicates a real and pronounced decline over time in the rate of adaptation, as opposed to an inability to detect further adaptation based on competitions with an increasingly distant ancestor.

B. Cell Size and Yield

One of those most conspicuous changes in the evolving *E. coli* populations is the average size of individual cells. Without knowing the direction of change in advance, one would probably guess that individual cells have become smaller. The evolved cells are growing and dividing much faster; and smaller cells have a higher surface-to-volume ratio, all else equal, which would be favorable for nutrient transport into the cells. In fact, however, the average cell size substantially increased over time in all 12 populations (Lenski and Travisano 1994; Vasi et al. 1994; Elena et al. 1996; Lenski and Mongold 2000). After 10,000 generations the grand mean cell volume was about twice that of the ancestor, as shown in Fig. 8.3.

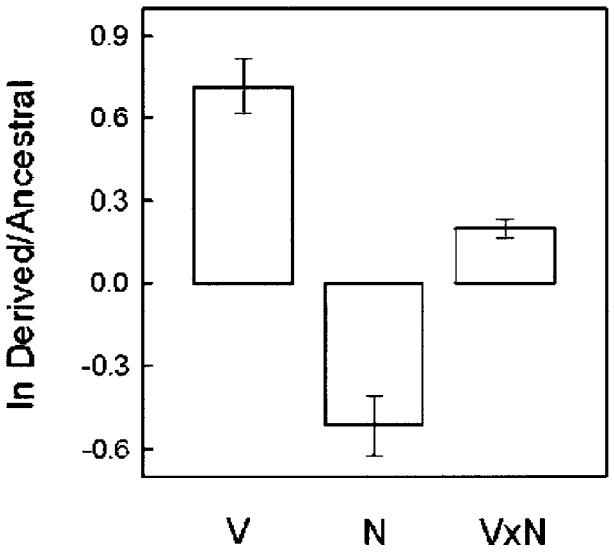

Fig. 8.3. Evolved changes after 10,000 generations in average cell volume (V), cell number at stationary phase (N), and biovolume yield (V×N). Each value is the mean over 12 populations of the ln-transformed ratio of evolved to ancestral values. Error bars are 95% confidence intervals. Source: Lenski and Mongold 2000. Used by permission of, and copyright 2000 by, Oxford University Press.

The trajectories for cell size and relative fitness are quite similar in their general form, in both cases changing much faster early in the experiment than late. Moreover, the step-like gains in fitness are temporally associated with corresponding changes in average cell size, implying that the mutations responsible for adaptation often caused a correlated increase in cell size (Elena et al. 1996). However, the evolved populations are more variable in their average cell volume than in their fitness (Lenski and Travisano 1994). Also, a few populations produce cells that are more spherical than the ancestor, although in most populations the overall rod-shape is similar to the ancestor (Lenski and Mongold 2000).

What can account for these unexpected changes in cell size and shape? We do not fully understand the relationship between the changes in cell morphology and fitness, although we can exclude several plausible explanations and thereby gain some understanding that may guide future work. Our initial cell-volume measurements were made on stationary-phase populations, when the bacteria have exhausted the glucose and ceased growth. Therefore, one plausible hypothesis for the larger cell size in the evolved populations is that the cell cycle has been altered such that cells stop completing their divisions as glucose becomes limiting and thereby produce more "doublets" at this stage. This hypothesis is rejected by the observation that the evolved cells are also larger in exponential-phase growth (Vasi et al. 1994) and, moreover, the evolved populations do not contain any more doublets at stationary phase (Lenski and Mongold 2000). A second plausible hypothesis is that the cells, while larger in volume, may not actually contain more biomass but instead might be larger sacks of water. Such an effect could result from some change in osmoregulation, for example. However, this hypothesis was rejected by measurements of the three main macromolecular constituents—protein, RNA, and DNA, which comprise >75% of the cell dry weight—that showed their concentrations are similar in the ancestral and evolved cells (Lenski et al. 1998).

It has been known for many years that any given strain of *E. coli* (ignoring evolution for the moment) produces larger cells when it is growing fast than when it is growing slow. This effect may seem counterintuitive but nonetheless it occurs, apparently because faster-growing cells are stuffed with more ribosomes and more copies of their replicating genome. (*E. coli* is genetically haploid, with a single circular chromosome, but an individual cell has several copies of its chromosome when it is growing under favorable conditions.) Hence, a third plausible hypothesis for the larger cell size of the evolved populations is that this change is simply an extension of the phenotypic correlation between growth rate and cell size into a genetic context. In other words, the long-term experiment

selected faster-growing cells, faster-growing cells tend to be larger, and thus their larger size is another manifestation of this phenotypic correlation. At first, it was not obvious to us how we could test this hypothesis, which requires disentangling the phenotypic and genetic correlations. However, it was possible to test it using a chemostat, which is a continuous-culture device that allows one to vary growth rate by changing flow rate through an open system. In particular, we could force evolved and ancestral clones to grow at the same rate in separate chemostats, and ask whether they achieved the same cell size, as this hypothesis would predict. We performed this experiment across a range of different growth rates and saw the expected positive phenotypic correlation between growth rate and cell size in both ancestral and evolved clones (Mongold and Lenski 1996). Nonetheless, the evolved cells were larger even when they grew at the same rate as the ancestor. This phenotypic correlation is thus consistent with, and contributes to, the increased size of the evolved cells. However, it is not sufficient to account for even half of the observed evolutionary change in cell size, based on the allometric relationship between growth rate and cell size measured for the ancestral genotype.

A fourth hypothesis is that we somehow selected inadvertently for larger cells. However, there is nothing in our procedures that would directly select cells on the basis of their size; for example, there is no filtration that could differentially retain larger cells, nor did we ever inspect cells visually as a basis for propagation.

A fifth possibility is that larger cells can acquire more resources. Although a larger cell has greater exposed surface than a small one of the same shape, the larger cell also has a reduced ratio of surface area to volume, which is allometrically unfavorable for resource acquisition in a constant environment. That is, a smaller cell that produces two daughter cells of the same small size should be able to acquire the necessary resources faster than a larger cell that produces two daughter cells of the same large size, all else being equal. However, the culture conditions in the long-term evolution experiment included a daily cycle of feast and famine as populations were diluted into fresh medium and then exhausted the glucose. It is possible that, in such fluctuating conditions, larger cells may have an advantage by virtue of their greater total surface area, which could allow them to acquire and sequester more glucose than they can immediately use, and then convert that surplus into progeny as the resources became depleted. Or perhaps larger cells, by virtue of having greater metabolic reserves, can respond more quickly to the sudden availability of resources and thereby commence growth more quickly (Lenski and Mongold 2000).

A sixth hypothesis is that the optimal cell size of *E. coli* in its natural environment is smaller than the optimum under the experimental regime, and hence the populations evolved larger cells. Unfortunately, the forces that shape cell size in either case are unknown, although a speculative scenario may serve to illustrate this hypothesis. In nature, *E. coli* is subject to attack by lytic viruses and other consumers, and one could imagine that size-selective predation favors cells that are smaller than would otherwise be optimal. But in our experimental regime, these consumers are absent, which would release the bacteria from size-selective predation and thereby favor larger cells. Let me reiterate that this scenario is only speculative, and it has not yet been tested in any way.

The three previous hypotheses imply selection on cell size per se. A seventh hypothesis is that the cell volume changes are merely a correlated response to some other more fundamental physiological change, despite the consistent pattern of increases in all 12 populations. (The third hypothesis, based on the physiological relationship between growth rate and cell size, is perhaps a specific example of this more general hypothesis.) For example, as will be described in Section II.D, the evolving populations have tended to become ecological specialists by down-regulating various catabolic pathways that are unnecessary in the experimental environment. Such changes are advantageous because they channel cellular metabolism to more productive pathways. Graña and Acerenza (2001) have recently produced a model of the bacterial cell cycle that suggests this reduction in wasteful expression can explain the observed evolution of larger cells. Their model fits nicely with several features of this system, although some independent test of its predictions would be desirable (for example, showing an evolutionary reduction in cell volume in a complex environment that favored expanded physiological capabilities). In any case, the model of Graña and Acerenza provides a plausible example of how changes in cell size may arise from selection on some general aspect of cell physiology and performance.

Whatever the reason (or combination of reasons) for the evolution of larger cells, this outcome appears to be closely connected with another change that also seems counterintuitive. The evolved populations yield fewer cells than the ancestor when they are grown separately under the standard conditions of the evolution experiment (Fig. 8.3). One might think that the evolved populations should produce more cells than the ancestor; after all, they have become more fit. Evolved cells do generally out-number the ancestors at the end of a competition assay, but not when they are grown separately. This difference has nothing to do with frequency-dependent selection or other complex interactions. It arises

because relative fitness does not measure the difference in numbers of organisms per se, but rather it measures the difference in the net rates of change in those numbers. All 12 evolved populations produce fewer, but larger, cells. The evolved cells each acquire and consume more of the limiting glucose, at a rate that is evidently disproportionate even to their larger volume. As a consequence, the larger evolved cells can exclude their smaller ancestors in the scramble competition for this limiting resource.

These two rather unexpected changes, larger cells that are fewer in number, combine to produce another change that is more readily understood. The product of average cell volume and numerical yield—that is, total biovolume produced—increased significantly in all 12 populations (Fig. 8.3). Coupled with the fact that concentrations of the main cell constituents remained nearly constant during the evolution experiment, this increased biovolume implies improved efficiency in the physiological conversion of the limiting glucose to biomass. Also, although the evolved populations are quite variable in both their average cell volume and numerical yield, these two traits vary inversely, such that total biovolume is almost constant (Lenski and Mongold 2000). This constancy suggests that the evolved populations have reached some limiting efficiency, but have done so by different routes.

C. Mutation Rate

Mutation is sometimes discussed as though it were an unavoidable consequence of copying mistakes during DNA replication. In fact, however, most organisms have exquisite molecular machinery that can identify and correct incipient errors, before they become mutations, using enzymes that proofread and repair DNA (Friedberg et al. 1995). When an organism's repair pathway malfunctions, such as by a mutation in one of the underlying genes, then its genomic mutation rate increases, often dramatically. [As an aside, it is interesting that many genes involved in DNA repair in bacteria have homologous genes in humans, and mutational defects in these homologous genes have been associated with certain cancers (Friedberg et al. 1995).]

One of the most interesting changes that has occurred in the long-term *E. coli* experiment is that several populations have evolved defects in their DNA repair (Sniegowski et al. 1997). These defects were demonstrated as follows. First, Luria-Delbrück fluctuation tests were run to measure the rate of spontaneous mutation at two or three genetic loci in which mutations produce a phenotypic change that is easily scored.

After 10,000 generations, three of the 12 populations exhibited greatly elevated mutation rates at each locus tested, whereas the other nine retained mutation rates similar to the ancestor, as shown in Fig. 8.4 for one locus. Across the loci tested, the increase in mutation rate in the "mutator" lines was about 100-fold. Second, such elevated mutation rates generally result from a loss of function, and thus genetic complementation tests were performed. Clones that possessed the mutator phenotype were made mero-diploid by introducing, one at a time, plasmids that encode functional copies of one of the genes involved in DNA repair. These tests revealed that the three hypermutable lines had become defective in the methyl-directed mismatch repair pathway, although in two different loci (Sniegowski et al. 1997). These genes were later sequenced to identify the precise mutations in the mutator lines (Shaver et al. 2002). Finally, clones isolated from earlier time points were tested to determine when hypermutability had evolved. In two populations, the mutator type became numerically dominant fairly early (around generations 2,500 and 3,000), and in the other population somewhat later (around 8,500 generations). All three lines that became mutators retained this phenotype through generation 20,000. A fourth population evolved a mutator phenotype after 15,000 generations (Cooper and Lenski 2000). The affected locus and other properties of this

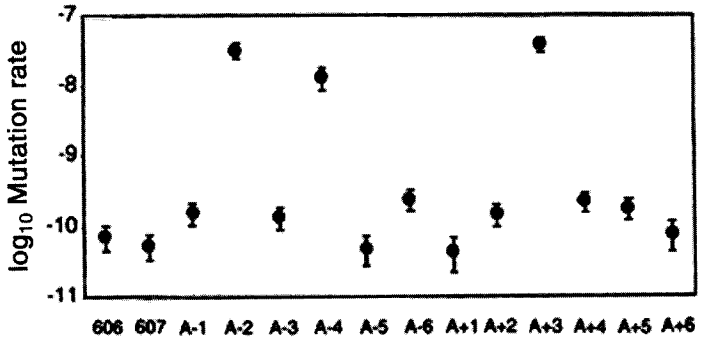

Fig. 8.4. Three of the 12 long-term *E. coli* populations evolved hypermutable phenotypes by 10,000 generations. These data show mutation rates for resistance to an antibiotic, nalidixic acid, measured in the ancestor (606 and 607) and evolved populations (A-1 through A+6). Error bars are approximate 95% confidence intervals. Notice that three evolved populations have rates that are about 100-fold higher than the ancestor or other evolved lines. Genetic analyses showed that these mutators had become defective in the methyl-directed mismatch repair pathway. Source: Sniegowski et al. 1997. Reprinted with permission of *Nature*.

late mutator have not yet been as thoroughly studied as the three that became mutators earlier.

The discovery that several populations evolved hypermutability raises some intriguing questions about both the causes and consequences of this phenomenon. How did the mutations that eliminated DNA repair become common? Their spread is puzzling because most mutations are deleterious; any mutation that increases the rate of other mutations should tend to increase genetic load and would thus be detrimental. Given that mutators did spread in some populations, but not in others, how does this difference affect subsequent phenotypic and genomic evolution? Do the mutator populations, by virtue of increased genetic variation, reach higher fitness than those populations that retained the ancestral mutation rate? And how does the increased mutation rate affect the rate of molecular evolution?

Let us begin by considering the consequences of elevated mutation rates, which will also give us some insight into understanding how the mutations that produced the mutator phenotype spread in their populations. We can assume, to a first approximation, that the rates of neutral, beneficial, and deleterious mutations are equally affected by the increase in mutation rate caused by defective DNA repair. Because the rate of substitution of *neutral* mutations depends only on the corresponding mutation rate (and not on population size or background selection), then we should clearly expect the lines that became mutators to accumulate many more neutral mutations than the other lines.

With respect to *beneficial* mutations, and the resulting rate of adaptation, the expectation is more complex. On the one hand, the waiting time for beneficial mutations to appear will be shorter if the mutation rate is higher. On the other hand, large populations may not spend much time waiting for a beneficial mutation, and the time required for a beneficial mutation to spread to fixation thus becomes important. As noted earlier, the most highly beneficial mutations in this experiment confer about a 10% selective advantage; such a mutation would (assuming it is lucky enough to avoid extinction by drift when it is still rare) require some 250 generations to increase from a single cell to the majority type (Lenski et al. 1991). Beneficial mutations that conferred a lesser advantage would take even longer to become the majority type. That long period spent as a minority, coupled with the clonal nature of the populations, means there is ample opportunity for a *more* beneficial mutation to occur on another background and ultimately out-compete the one that occurred first. As a consequence of this clonal interference, the rate of adaptation in a large asexual population shows diminishing returns

with an increasing supply rate of beneficial mutations. This effect has been confirmed both theoretically (Gerrish and Lenski 1998) and in a separate experiment with bacteria in which mutation rate and population size were manipulated (De Visser et al. 1999). These studies indicate that, for the parameters most relevant to the long-term experiment, the increase in the rate of genetic adaptation caused by a 100-fold increase in mutation rate should be rather small. Consistent with these expectations, the populations that became mutators evolved only slightly higher fitness than the other populations, and even this difference is not statistically significant. The important point here is that any acceleration in the rate of adaptation is far less than proportionate to the increase in mutation rate. Hence, the increase in mutation rate has very different consequences for the substitution rates of neutral and beneficial mutations.

Finally, one would not expect *deleterious* mutations to be substituted by pure drift in such large populations as studied here, unless a mutation has a truly negligible effect. However, a slightly deleterious mutation could be substituted by hitchhiking with a beneficial mutation that occurred in the same background. But, as explained in the previous paragraph, the mutator lines do not undergo many more beneficial substitutions than the other lines. Thus, there are not many more chances for deleterious mutations to hitchhike to fixation in the mutator populations. More important than the substitution of deleterious mutations, however, is the effect of hypermutability on the genetic load resulting from deleterious mutations that have not been fixed. This load will reduce the mean fitness of the mutator lines (a consequence) as well as impede the substitution of mutations that cause the mutator phenotype (a cause that will be examined below). In a haploid population, the theoretical genetic load caused by deleterious mutations is approximately equal to the genomic rate of deleterious mutation, and thus should increase proportionately with the mutation rate (Sniegowski et al. 2000). But because the total genome-wide rate of deleterious mutations was quite low in the ancestor, the load associated with a 100-fold increase in mutation rate is still not very large. The genome-wide mutation rate in repair-proficient *E. coli* is about 0.0025 mutations per genome per generation (Drake et al. 1998), perhaps even lower (Lenski et al. 2003). This includes neutral mutations, and the rate of deleterious mutation estimated from mutation-accumulation experiments is still lower, only about 0.0002 per genome per generation (Kibota and Lynch 1996). Even a 100-fold increase in this rate would produce a genetic load of only 2%. Hence, the mutators do not suffer a huge reduction in mean fitness

associated with an increased genetic load, nor will selection against deleterious mutations be sufficiently strong to impede the spread of mutator alleles that are linked to beneficial mutations of large effect.

Having examined the likely consequences of the hypermutability that evolved in some of the long-term *E. coli* populations, let us return to the question of how a mutation that produced this effect could be substituted given its increased load of deleterious mutations. First, there is no evidence that loss of DNA repair confers any direct fitness advantage (Chao and Cox 1983; Shaver et al. 2002). This result is based on direct competitions between strains that differ only by a mutation in a DNA repair gene, and the small fitness effect that is measured is consistent with the slight increase in genetic load calculated in the preceding paragraph. Second, the idea that hypermutability might be favored because it increases evolvability fails in several respects. Evolution is not a goal-directed process. Group selection may sometimes give rise to features that appear goal-directed, but the large population size and lack of spatial structure in the long-term experiment (populations were constantly dispersed by shaking the liquid medium) provide no meaningful opportunity for group selection favoring greater evolvability to overcome even a 2% higher load of deleterious mutations. What can explain the substitution of mutations causing defects in DNA repair?

The answer appears to rely on hitchhiking, but with an important twist. The mutator allele is deleterious, but it can hitchhike with a beneficial mutation that more than offsets its disadvantage. That statement is true for any deleterious mutation, especially one that causes little harm. The twist is that the mutator allele is much more likely to generate a beneficial mutation in its own background than is the wild-type allele that encodes functional DNA repair. In essence, the mutator promotes its own hitchhiking (Taddei et al. 1997).

Consider the following "back-of-the-envelope" calculations. There are six or so genes in *E. coli* that encode DNA repair functions which, if eliminated individually, would cause an increase in the mutation rate of similar magnitude to what evolved in several of the long-term lines. Together these genes comprise about 12,000 base-pairs that are at-risk for mutation and perhaps one-third of them would, if mutated, disrupt DNA repair. If the base-pair mutation rate in repair-proficient *E. coli* is about 5×10^{-10} per generation, then the overall rate of mutations to produce a mutator is perhaps 2×10^{-6} per generation. In fact, one of the genes involved in DNA repair contains a repeated motif that appears prone to mutations (Shaver et al. 2002), and this rate may therefore be an underestimate. Given this rate of production and a 2% fitness cost of

being a mutator, then the equilibrium frequency of mutators in a haploid population, under mutation-selection balance, should be roughly 1×10^{-4}. This mutator sub-population has a 100-fold higher rate, per capita, of generating the next beneficial mutation that eventually goes to fixation, which would bring the mutator allele along as a hitchhiker. Therefore, the mutator allele should have a probability of about 1% of being substituted along with the next beneficial mutation. That may not seem like much, but remember that each population substituted perhaps 10 or 20 beneficial mutations. Any one of these, at least among those providing a benefit much greater than 2%, would have allowed the mutator sub-population a chance to generate the next winner and then hitchhike along with it. The fact that 4 of the 12 populations became mutators is reasonably consistent with this scenario, especially given the uncertainties associated with exact values.

The main conclusions are paradoxical. Mutators are associated with rapid adaptation, but not because they accelerate genetic adaptation to any great extent in a large population. Instead, rapid adaptation, such as occurs in a new environment, provides more opportunities for a mutator to hitchhike with a beneficial mutation, which the mutator subpopulation may produce just a bit sooner than would otherwise occur. Over a long time, one could imagine this process reversing itself. That is, the populations may eventually be so well adapted to the experimental regime that the cost of deleterious mutations exceeds the benefit of any remaining advantageous mutation. At that point, a mutation that restored DNA repair function would provide a selective advantage, and there would be no beneficial mutations in other clones to interfere with its spread. However, the advantage of this anti-mutator allele would be proportional to the reduction in genetic load, about 2%, and hence its spread would be fairly slow. Most other accessible beneficial mutations conferring a greater advantage would have to be substituted in the population before this process of reversal could effectively commence. Hence, the mutators that have become substituted will probably be difficult to displace.

D. Ecological Specialization

It is clear from the fitness trajectories that the evolved *E. coli* lines have substantially improved their performance in the selective environment. But what happened to their performance in other environments as they adapted to this particular set of conditions? Have they become highly specialized, or do they retain their ancestral performance levels in

other environments? If the bacteria did become more specialized, was this a consequence of pleiotropic tradeoffs associated with the mutations that adapted them to the experimental environment? Or was specialization a consequence of decay by random drift of mutations in genes encoding physiological functions that were unimportant in the selective environment?

The first set of experiments that we performed to address these questions used bacterial clones sampled at generation 2,000 (before any of the populations had become hypermutable). We tested the fitness of these clones relative to the ancestor in a series of environments that were identical in every respect to the selective environment, except that the sole carbon and energy source, glucose, was replaced by one of 11 other substrates (Travisano et al. 1995; Travisano and Lenski 1996). The evolved lines exhibited significant correlated declines in performance on two substrates, and significant correlated gains on five substrates. Interestingly, four of the substrates that showed correlated gains were among the five substrates that share with glucose their route of transport into the cell through both inner and outer membranes. In addition to these differences between substrates in average correlated response, the among-line genetic variance component also differed between substrates. The among-line variance was lower in glucose than in any of the other 11 substrates, indicating that the direct response to selection was more parallel than the correlated responses. There was also a tendency for the correlated responses to be more uniform for the substrates that used the same basic mode of transport as glucose, with greater divergence among populations in their responses to substrates that entered the cell by other pathways.

The second set of experiments to address this general set of issues differed in approach in several ways (Cooper and Lenski 2000). First, the evolutionary time frame extended to 20,000 generations. Second, several time points were examined to allow the trajectory of correlated responses to be examined. Third, many more growth substrates were tested using 96-well plates. Finally, to accommodate the plate format as well as the larger number of samples and substrates, measurements were made of growth kinetics as opposed to competitive fitness per se. These data, summarized in Fig. 8.5, showed a significant trend toward ecological specialization, with "catabolic breadth" (a measure of overall performance on 64 informative substrates other than glucose) declining by 42%, on average, after 20,000 generations. Almost all of the performance losses were quantitative, not qualitative (absolute), with one conspicuous exception: all 12 lines completely lost the ability to grow on D-ribose, as will be discussed in the next section.

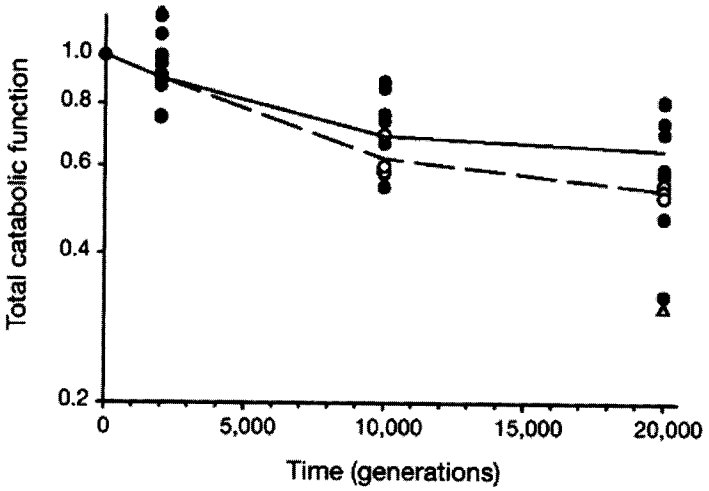

Fig. 8.5. Ecological specialization of the evolving populations indicated by declining catabolic breadth. The populations evolved in a minimal medium that contained glucose as the sole source of carbon and energy. Total catabolic function is a weighted average performance on 64 other substrates, standardized to an ancestral value of 1 and shown on a ln-transformed scale. Each point indicates one of the 12 populations. Closed circles represent populations that retained the ancestral mutation rate; open circles became mutators between 2,000 and 10,000 generations; and the open triangle became a mutator between 10,000 and 20,000 generations. The solid line indicates the mean of the nonmutator populations; the dashed line is the mean of the mutator populations excluding the one that evolved latest. Notice that the rate of decline decelerates in parallel to the rate of fitness gain (Fig. 8.1). Note also that the decline in total catabolic function is slightly faster in the populations that became mutators, but not nearly to the 100-fold extent that their mutation rates increased (Fig. 8.4). Source: Cooper and Lenski 2000. Reprinted with permission of *Nature*.

In principle, two distinct population-genetic processes could account for the trend toward specialization: antagonistic pleiotropy (AP) and mutation accumulation (MA). According to AP, the losses of performance on other resources result from tradeoffs, in which the same mutations that are beneficial in the glucose environment have detrimental effects in other environments. According to MA, the correlated losses of adaptation to the other environments were caused by mutations that drifted (or hitchhiked) to fixation but did not themselves contribute to adaptation to glucose.

Several lines of evidence indicate that AP was more important than MA for specialization that evolved in the long-term experiment (Cooper and Lenski 2000). First, the decay in catabolic breadth was faster early

in the experiment than late. This trajectory mirrors genetic adaptation, which was also more pronounced early than late, and hence supports AP. Second, the decay of catabolic breadth was not significantly greater in those lines that became hypermutable than in those that did not. The 100-fold higher mutation rate in the mutator lines should have led to a corresponding increase in MA and associated specialization by drift decay. The fact that specialization was not dramatically more pronounced in these mutator lines implies that the contribution of MA is secondary to that of AP. Third, much of the observed decay in catabolic breadth was concentrated in a subset of functions, a pattern more consistent with AP than with MA. Fourth, a separate study was performed by another group of losses of catabolic function in *E. coli* populations that were all mutators and, moreover, were subjected to severe bottlenecks (Funchain et al. 2000). Such bottlenecks promote MA by magnifying drift, which prevents the elimination of deleterious mutations and precludes the genetic adaptation that underlies AP. This other study observed no deceleration in the decay rate of catabolic breadth, nor was there much concentration of losses in a subset of catabolic functions. These differences in the trajectory and the pattern of specialization support the importance of AP in our long-term selection experiment. Fifth, as described in the next section, at least some of the mutations that contributed to glucose adaptation contributed to the decay of catabolic breadth, providing a direct demonstration of AP.

In summary, the extensive and largely parallel fitness gains in the glucose environment were accompanied by correlated changes in performance in other resource environments. These correlated responses were more divergent across the replicate lines than was the directly selected performance on glucose, and overall catabolic breadth tended to decline. The decay of catabolic breadth, and the resulting ecological specialization, are more consistent with the effects of antagonistic pleiotropy than with drift decay by mutation accumulation.

E. Ribose Catabolic Function

While performing the analyses of catabolic breadth, we discovered that all 12 populations had lost the ability to grow on D-ribose as a sole carbon source (Cooper and Lenski 2000). When we examined frozen samples to document the time course of these losses, we saw that the lines had invariably lost their ability to use ribose very quickly: seven in the first 500 generations, and all 12 of them by generation 2,000 (Cooper et al. 2001). Finding these rapid and parallel changes immediately suggested that the loss of the ribose catabolic function was highly advanta-

geous in the glucose environment. We therefore set out to discover the genetic bases of these mutations, and to measure their fitness effects independent of other genetic changes that occurred during the evolution experiment.

Early in this work, we made another discovery that surprised us: mutations that cause the loss of the ribose utilization function occur at a surprisingly high rate, about 5×10^{-5} per cell generation based on Luria-Delbrück fluctuation tests, even in a strain with functional DNA repair (Cooper et al. 2001). This rate is more than an order of magnitude higher than what one would expect for point mutations for the entire *rbs* operon. Molecular genetic analyses revealed that the losses of the ribose catabolic function in the evolved lines involved deletions of part or all of the *rbs* operon, as illustrated in Fig. 8.6. One endpoint of the deletion differed in every case, but the other deletion endpoint was precisely the same in all cases, and coincided exactly with the edge of a mobile element, called IS*150*, that happened to be immediately adjacent to the *rbs* operon in our ancestral *E. coli* strain. Evidently, this element was responsible for the genetic instability of the ribose catabolic function.

This localized hypermutability led us to question our initial assumption that such striking parallelism was caused by strong selection. Perhaps instead the parallel losses merely reflected the genetic instability of the *rbs* operon. To examine this issue further, we isolated several *rbs*-deletion mutations on the ancestral background, and we also engineered a deletion genotype not involving the nearby IS*150* element. When each of these Rbs^- genotypes was put in competition with the otherwise identical Rbs^+ ancestor, in the same glucose medium as used in the long-term experiment, the Rbs^- competitor had a fitness advantage of about 1–2% (Cooper et al. 2001). Thus, positive selection for loss of the ribose catabolic function and its genetic instability both contributed to the rapid and parallel losses. A mathematical analysis of the temporal spread of the Rbs^- genotypes, using the measured mutation rate and selection coefficient, showed that neither the hypermutability of that locus nor a 2% selective advantage alone was sufficient to account for the substitution of the Rbs^- mutants in only 2,000 generations. However, the two processes combined could explain the rapid evolutionary loss of the ribose-catabolic function.

F. More and More Genetics

The preceding section gave a concrete example of the type of research that we are most actively pursuing with the long-term *E. coli* populations. We seek not only to find mutations that were substituted in these

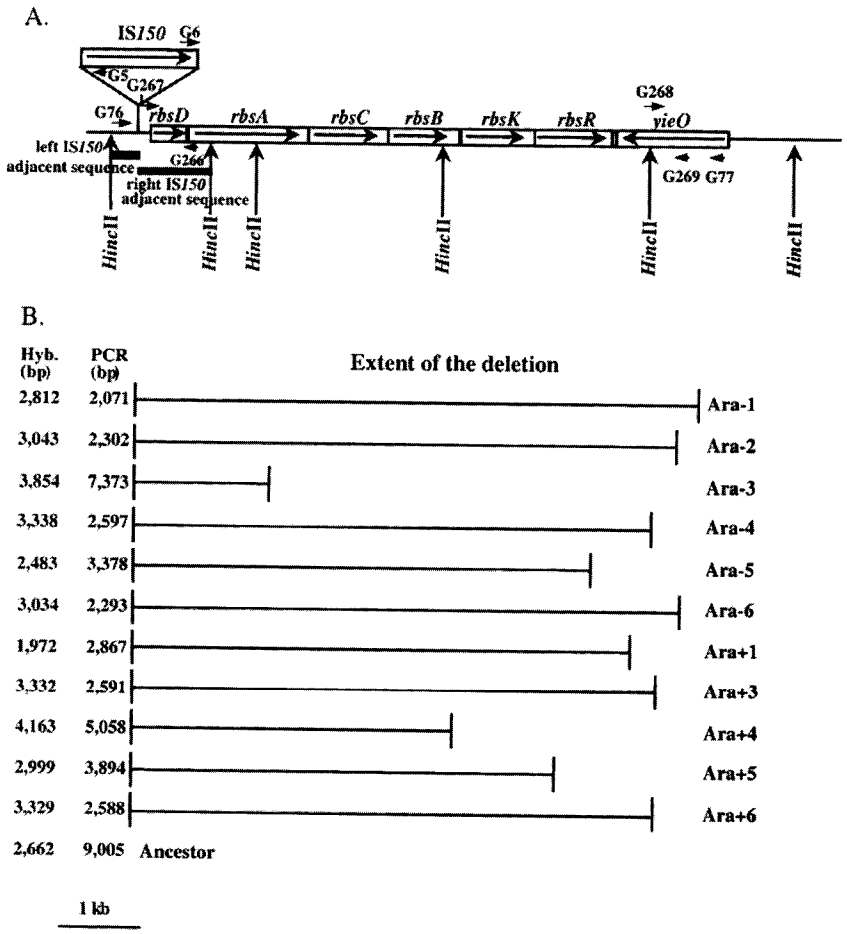

Fig. 8.6. Deletions in the evolving populations of the *rbs* operon led to parallel losses of their ability to grow on D-ribose. A. Map of the *rbs* operon shows an IS*150* element located upstream of the first gene in the ancestral strain. The boxes denote genes, and arrows within them show the direction of transcription. B. Physical extent of deletion mutations in 11 of the populations. All of the deletions have the IS*150* element as their left endpoint, but they have different right endpoints. Source: Cooper et al. 2001. Reprinted with permission of American Society for Microbiology.

populations, but we also want to understand how particular mutations spread and what other traits they influence. To do so, we must reconstruct genotypes that differ only by the mutation at hand, in order to isolate its effects from other mutations that were also substituted in the same evolving line. Are most of the mutations that were substituted neu-

tral or beneficial? Are beneficial mutations found in the same genes and pathways in many or all of the lines, or did each population discover a unique way of adapting to the selective environment? Can we integrate existing knowledge of the genetics and physiology of *E. coli* with the various findings of our long-term evolution experiment? Our work on these fronts is well underway, and we have many interesting results, but some are not yet published. Hence, in this section, I will sometimes speak in general terms about the approaches that we are using to discover mutations. I hope that readers who are interested enough to have read thus far into my review will also be interested in following our genetic discoveries and analyses of mutational effects as they are published over the coming years.

1. IS-Mediated Mutations. Insertion sequences, or IS elements, are mobile elements present in most bacterial genomes. IS elements encode functions that promote their own transposition, and hence they are mutagenic. Through recombination between homologous elements, IS elements also promote rearrangements, including deletions and inversions. The mutations caused by IS elements can be discovered much more readily than point mutations, because IS mutations cause discernible changes in the size of restriction fragments and because knowledge of their sequence allows one to make probes and primers for finding their genomic locations.

Two of the lines were chosen as focal material for investigating IS-mediated mutations. Many clones from several time-points through generation 10,000 were used as study material, and numerous IS-mediated mutations were indeed discovered (Papadopoulos et al. 1999), as shown in Fig. 8.7. The resulting genetic diversity was impressive: in one population all 11 clones tested at generation 10,000 had distinct genotypes based on IS elements, while in the other line testing of 13 clones revealed 10 different genotypes. Given all this variation, we chose to focus on those IS-mediated mutations that were substituted in one of these two populations in this period, reasoning that such mutations were more likely to be beneficial and hence of greater interest. In addition to the IS-mediated *rbs* deletions discussed in Section II.E, nine other IS substitutions were genetically characterized in the two focal populations (Schneider et al. 2000). These included six new insertions, two large inversions, and a deletion. The new insertions are of particular interest because they involve interesting genes, and because such mutations are easier to work with than inversions or deletions. We are presently performing studies in which we construct genotypes that are identical except for a mutation in the affected gene, in order to measure the mutation's effects on fitness

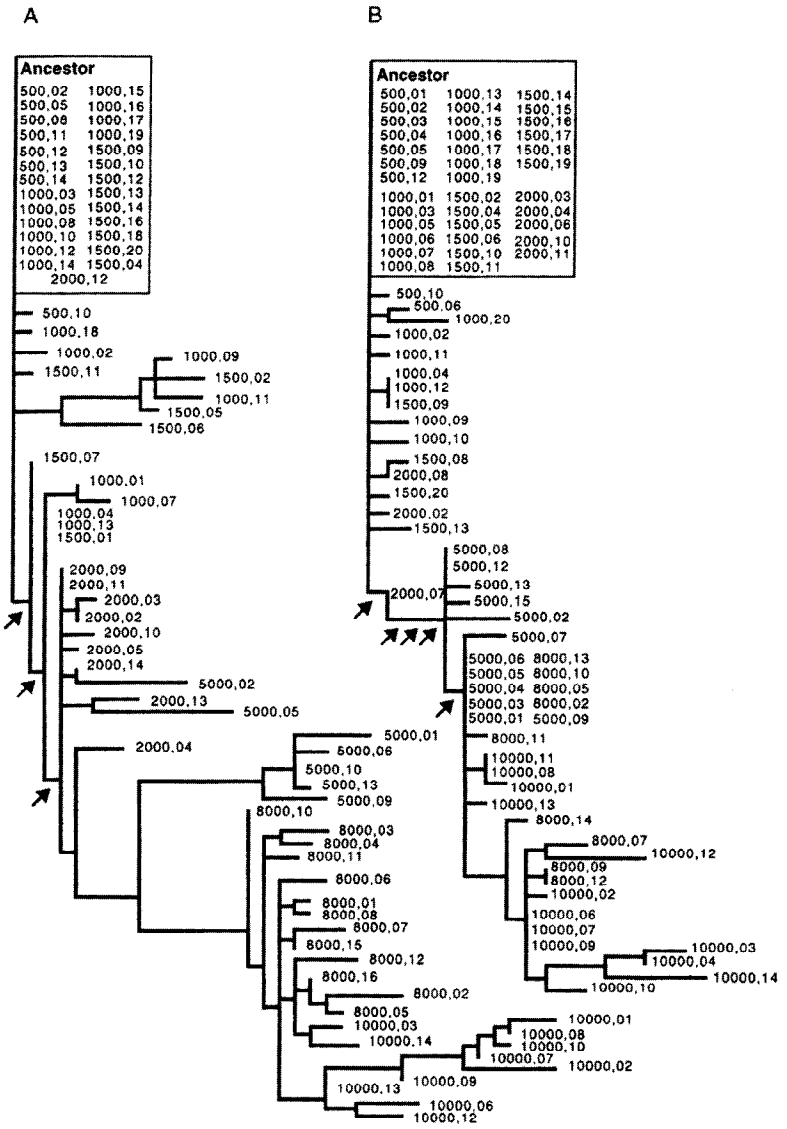

Fig. 8.7. Evolving DNA "fingerprints" in two of the populations, obtained using IS elements as probes. A and B show phylogenies constructed for two populations through 10,000 generations. The boxes represent the ancestral clone and clones sampled in later generations that were indistinguishable from the ancestor by this approach. Each number indicates a clone, with the first part showing generation number and the second an arbitrary designation for clones sampled in the same generation. Arrows indicate some of the IS-mediated mutations that were substituted in the population. Source: Papadopoulos et al. 1999.

and other traits. If a particular IS-mediated mutation proves to be beneficial, then the affected gene becomes an interesting candidate for study in the other 11 lines to determine, for example, whether they also substituted similar mutations.

2. Transposon Tagging. Another approach to finding beneficial mutations analyzes the linkage between fitness-enhancing mutations and genetic markers. These markers must be deliberately introduced into the bacterial chromosome for this purpose, owing to the initial homogeneity of the material. This approach is elegant, but also laborious. Mark Stanek and I (ms. in prep.) used this approach to find a beneficial mutation as follows. Starting with a clone from one population at generation 10,000, we produced a library of more than 1,000 marked clones. Each clone carried a transposon that randomly inserted at a different genomic location, and that encodes a marker that can be readily selected. A viral transduction system was then used to move bits of DNA from the pooled library into the ancestral genotype, and selection was applied to the marker (to eliminate non-recombinants and other uninformative genotypes). We then had a pool of recombinant genotypes that carried bits of the evolved clone's genome, including the marker, in the ancestral background.

The challenge was to find an interesting "needle in a haystack," which we did as follows. First, we propagated the recombinant pool for about 40 generations, too little to allow much evolution de novo, but enough to enrich slightly those recombinants that acquired a beneficial mutation from the evolved clone. Second, we used the fact that cell size and fitness were strongly correlated during evolution, and the much greater ease of screening cell size than fitness, to isolate promising recombinant clones for further study. Third, after finding a recombinant clone that produced much larger cells than the ancestor, the two were placed in competition to confirm that the recombinant was indeed more fit. Fourth, having in hand a recombinant clone containing a beneficial mutation and a nearby marker, we more precisely quantified the linkage between the two by co-transduction. An analysis of these data indicated that the beneficial mutation of interest was approximately 3,000 base-pairs from the transposon that carried the marker. Fifth, we sequenced about 5,000 base-pairs in each direction from the transposon, and we found one and only one difference between the ancestral and recombinant clones. This mutation is therefore responsible for the differences in both fitness and cell size between these clones. Finally, we went back into the stored samples to determine the trajectory of that allele in the population in which it arose, and whether the same or similar mutations were substituted in other lines.

3. Expression Arrays. Section II.D described how we used 96-well plates to screen changes in many catabolic functions, which led us to identify deletion mutations in the *rbs* operon described in Section II.E. Another data-intensive approach we have recently pursued uses DNA arrays, which allow one to measure simultaneously the changes in expression for almost all the genes in the *E. coli* genome (Cooper et al. 2003). In brief, one harvests the mRNA from a cell, makes the corresponding cDNA that is radio-labeled, and hybridizes that cDNA to a membrane-bound array of all 4,290 of the organism's open-reading frames. We then compared the patterns of gene expression of the ancestor and evolved lines when they had been separately grown under identical conditions. To minimize the vexing statistical problem of avoiding false positives in such large data sets (with >4,000 genes and a 0.05 significance level, one could easily get >200 false positives), we have so far focused our attention on changes in gene expression that occurred in two independently evolved lines, which again serve as focal material, over 20,000 generations.

The overall gene-expression profiles of both the evolved lines were significantly more divergent from the ancestor than were controls, in which the genetically marked variants of the ancestor were compared to one another. The two independently evolved lines were also more divergent from one another than were the ancestral controls. If the two evolved lines had changed their patterns of gene expression in completely different ways, we would have expected them to be about twice as divergent from one another as they were on average from the ancestor. In fact, however, they were less divergent from one another than either was from the ancestor. This finding indicates strong (though not complete) parallel evolution of gene expression profiles.

We then identified 59 genes whose individual expression levels had changed significantly in both independently evolved lines. Remarkably, all 59 changed in the same direction in both lines (both lines with increased expression, or both with decreased expression), even though this concordance was not part of the statistical test used to find these 59 genes. At this point, it is very important to emphasize that these expression data are phenotypic data; they are not genetic data, despite the nucleic-acid based methods employed. For example, a single mutation in some regulatory gene could cause tens or more of coordinated changes in expression of other genes. By using the abundant knowledge of metabolic and regulatory pathways in *E. coli*, we inspected the parallel patterns of changes in gene expression in order to identify some candidate genes for sequencing and, if mutations were found, manipulation.

To make a long and complex story short and simple, we identified a mutation in one of the two focal lines in a gene called *spoT* (Cooper et

al. 2003). The encoded SpoT protein is a bi-functional protein that can add and remove a phosphate moiety from a molecule called ppGpp, an important effector whose intracellular concentration has cascading effects on gene expression. We then moved the *spoT* mutation we had discovered into the ancestral genetic background and showed that it conferred a competitive fitness advantage of almost 10%. Moreover, moving the evolved *spoT* allele into the ancestor produced many of the same changes in gene expression that we had used to discover that mutation. Interestingly, there was no *spoT* mutation in the other focal line, and some other (presently unknown) mutation evidently has similar effects on gene expression. Moreover, this unknown mutation evidently has similar effects on fitness because moving the evolved *spoT* mutation from one line to the other conferred no fitness advantage. However, sequencing revealed non-synonymous substitutions in *spoT* in seven of the other ten independently evolved populations. As we will see in the next section, finding mutations in the same gene in 8 of 12 lines is quite unlike what is seen in most of the genome, providing further evidence for the adaptive significance of these substitutions.

4. Random Sequencing. IS-mediated mutations can be found by screening the entire genomes of ancestral and evolved clones, because these mutations produce large changes in the size of restriction fragments that can be readily detected. By contrast, it is much more difficult to detect point mutations or other small mutations. Of course, it is possible now to sequence an entire bacterial genome, but at present the costs are too great to do this for all the populations, clones within populations, and sample time points that would interest us. Therefore, as a first step in this direction, we sequenced 36 randomly chosen gene regions in 50 clones, including the two variants of the ancestor and two clones from two time points for all 12 populations (Lenski et al. 2003). We found a total of only 10 mutations in these randomly chosen regions; all were found in populations that became mutators and, moreover, all had sequence signatures that were typical of mutator genotypes. Based on the subset of these mutations that were synonymous (hence presumably neutral) and had been substituted in their populations, and using independent data on the relative mutation rates of the mutator and repair-proficient states, we estimated the ancestral mutation rate as about 1.4×10^{-10} per base-pair per generation. This value is somewhat lower than a widely cited estimate based on genetic experiments (Drake et al. 1998), but it is higher than another estimate calculated from inter-specific sequence comparisons (Ochman et al. 1999).

These data from randomly chosen genes also provide a useful control for understanding how much background variation to expect in candidate genes where we have some phenotypic basis for sequencing them, such as the *spoT* candidate deduced from studying expression arrays. In none of the 36 randomly chosen genes did we find mutations in even two of the 12 evolved lines (Lenski et al. 2003), whereas eight of these lines had non-synonymous substitutions in *spoT* (Cooper et al. 2003).

In addition to sequencing random genes, Christine Borland and I are currently exploring methods that might allow more rapid discovery of single-nucleotide polymorphisms, or SNPs. Such methods are based on hybridizing fragments of genomic DNA derived from two different clones, followed by molecular methods to detect mismatches between the fragments. Perhaps in a few years we will also pursue complete genome sequences for our ancestral strain and several evolved clones.

III. CONCLUSIONS

For over a decade, my group has maintained 12 populations of the bacterium *Escherichia coli*, founded from the same ancestor, in a simple defined laboratory environment in which glucose provides the sole source of carbon and energy. The evolving populations have undergone more than 20,000 generations of binary fission. Each population had a billion or so mutations appear, although only a tiny fraction were actually substituted in the population.

Extensive phenotypic evolution has occurred, including substantial gains in competitive fitness. After 20,000 generations, the evolved bacteria on average grow about 70% faster than the ancestor when they compete in the same environment. Average cell size also dramatically increased in the evolving populations. The evolved bacteria have tended to become glucose specialists, with mostly subtle reductions in many other catabolic functions. This specialization reflects pleiotropic tradeoffs of mutations that are beneficial in glucose medium, much more than it does drift accumulation of neutral mutations in unused genes. Several populations evolved defects in DNA repair and became hypermutable as a consequence, but the mutator lines do not exhibit proportionately greater adaptation or specialization.

The rate of phenotypic evolution has decelerated over time. Deceleration in the rate of adaptation reflects each population's approach to an adaptive peak or plateau, not an inability to measure further adaptation owing to non-transitive competitive interactions. Evolution was often parallel across the replicate populations, not only for fitness in the selec-

tive environment but also for cell morphology and underlying physiology as reflected by performance on other substrates and patterns of gene expression. On the one hand, such parallelism is surprising given that the populations relied entirely on new mutations for genetic adaptation. On the other hand, the large population size and large number of generations ensured that almost any one-step mutation was accessible in every population.

We are now in the midst of finding mutations that have been substituted in these evolving populations. Several different approaches are being used, including sequencing both candidate and random genes. Once a mutation of interest is found, the gene must be manipulated, such as by moving the mutation onto an otherwise isogenic background, to test its effects on fitness and other traits of interest. For example, in the course of screening catabolic breadth, we found that all 12 populations had lost their ability to grow on ribose. Molecular genetic analysis found similar, although not identical, deletions of the *rbs* operon associated with an adjacent mobile genetic element. Both spontaneous and engineered mutations that eliminated this operon were produced in the ancestral background, and these mutations were shown to confer a selective advantage in the environment used in the long-term experiment. Similarly, evolved changes in gene-expression profiles led to the discovery of point mutations in the *spoT* gene in 8 of the 12 populations. By precisely moving an evolved *spoT* allele into the ancestral genome, we showed that this mutation provided a significant competitive advantage in the environment of the long-term evolution experiment.

LITERATURE CITED

Blattner, F. R., G. Plunkett, C. A. Bloch, N. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–1462.

Chao, L., and E. C. Cox. 1983. Competition between high and low mutating strains of *Escherichia coli*. Evolution 37:125–134.

Cooper, T. F., D. E. Rozen, and R. E. Lenski. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *E. 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 *E. coli* populations. Nature 407:736–739.

Cooper, V. S., D. Schneider, M. Blot, and R. E. Lenski. 2001. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *E. coli* B. J. Bacteriol. 183:2834–2841.

Coyne, J. A., and B. Charlesworth. 1996. Mechanisms of punctuated evolution. Science 274:1748–1749.

De Visser, J. A. G. M., and R. E. Lenski. 2002. Long-term experimental evolution in *Escherichia coli*. XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. BMC Evol. Biol. 2:19.

De Visser, J. A. G. M., C. W. Zeyl, P. J. Gerrish, J. L. Blanchard, and R. E. Lenski. 1999. Diminishing returns from mutation supply rate in asexual populations. Science 283:404–406.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. 1998. Rates of spontaneous mutation. Genetics 148:1667–1686.

Elena, S. F., and R. E. Lenski. 1997. Long-term experimental evolution in *Escherichia coli*. VII. Mechanisms maintaining genetic variability within populations. Evolution 51:1058–1067.

Elena, S. F., V. S. Cooper, and R. E. Lenski. 1996. Punctuated evolution caused by selection of rare beneficial mutations. Science 272:1802–1804.

Friedberg, E. C., G. C. Walker, and W. Siede. 1995. DNA repair and mutagenesis. ASM Press, Washington.

Funchain, P., A. Yeung, J. L. Stewart, R. Lin, M. M. Slupska, and J. H. Miller. 2000. The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. Genetics 154:959–970.

Gerrish, P. J., and R. E. Lenski. 1998. The fate of competing beneficial mutations in an asexual population. Genetica 102/103:127–144.

Gould, S. J. 1989. Wonderful life: the Burgess shale and the nature of history. Norton, New York.

Gould, S. J. 2002. The structure of evolutionary theory. Harvard Univ. Press, Cambridge.

Gould, S. J., and N. Eldredge. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. Paleobiology 3:115–151.

Graña, M., and L. Acerenza. 2001. A model combining cell physiology and population genetics to explain *Escherichia coli* experimental evolution. BMC Evol. Biol. 1:12.

Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection. V. Selection and mutation. Proc. Camb. Phil. Soc. 23:838–844.

Kibota, T. T., and M. Lynch. 1996. Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature 381:694–696.

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

Lenski, R. E., and J. A. Mongold. 2000. Cell size, shape, and fitness in evolving populations of bacteria. p. 221–235. In: J. H. Brown and G. B. West (eds.), Scaling in biology. Oxford Univ. Press, Oxford.

Lenski, R. E., and M. Travisano. 1994. Dynamics of adaptation and diversification: 10,000-generation experiment with bacterial populations. Proc. Natl. Acad. Sci. (USA) 91:6808–6814.

Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am. Nat. 138:1315–1341.

Lenski, R. E., J. A. Mongold, P. D. Sniegowski, M. Travisano, F. Vasi, P. J. Gerrish, and T. Schmidt. 1998. Evolution of competitive fitness in experimental populations of *E. coli*: what makes one genotype a better competitor than another? Antonie van Leeuwenhoek 73:35–47.

Lenski, R. E., C. L. Winkworth, and M. A. Riley. 2003. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. J. Mol. Evol. 56:498–508.

Mongold, J. A., and R. E. Lenski. 1996. Experimental rejection of a nonadaptive explanation for increased cell size in *Escherichia coli*. J. Bacteriol. 178:5333–5334.

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

Neidhardt, F. C., R. Curtiss, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (eds.). 1996. *Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press, Washington.

Ochman, H., S. Elwyn, and N. A. Moran. 1999. Calibrating bacterial evolution. Proc. Natl. Acad. Sci. (USA) 96:12638–12643.

Orr, H. A. 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution 52:935–949.

Papadopoulos, D., D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, and M. Blot. 1999. Genomic evolution during a 10,000-generation experiment with bacteria. Proc. Natl. Acad. Sci. (USA) 96:3807–3812.

Paquin, C., and J. Adams. 1982. Relative fitness can decrease in evolving populations of *S. cerevisiae*. Nature 306:368–371.

Rozen, D. E., and R. E. Lenski. 2000. Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. Am. Nat. 155:24–35.

Schneider, D., E. Duperchy, E. Coursange, R. E. Lenski, and M. Blot. 2000. Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements. Genetics 156:477–488.

Shaver, A. C., P. G. Dombrowski, J. Y. Sweeney, T. Treis, R. M. Zappala, and P. D. Sniegowski. 2002. Fitness evolution and the rise of mutator alleles in experimental *Escherichia coli* populations. Genetics 162:557–566.

Sniegowski, P. D., P. J. Gerrish, and R. E. Lenski. 1997. Evolution of high mutation rates in experimental populations of *Escherichia coli*. Nature 387:703–705.

Sniegowski, P. D., P. J. Gerrish, T. Johnson, and A. Shaver. 2000. Evolution of mutation rates: separating causes from consequences. BioEssays 22:1057–1066.

Souza, V., P. E. Turner, and R. E. Lenski. 1997. Long-term experimental evolution in *Escherichia coli*. V. Effects of recombination with immigrant genotypes on the rate of bacterial evolution. J. Evol. Biol. 10:743–769.

Taddei, F., M. Radman, J. Maynard Smith, B. Toupance, P.-H. Gouyon, and B. Godelle. 1997. Role of mutator alleles in adaptive evolution. Nature 387:700–703.

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.