J Mol Evol (2003) 56:498–508
DOI: 10.1007/s00239-002-2423-0

JOURNAL OF MOLECULAR EVOLUTION

Rates of DNA Sequence Evolution in Experimental Populations of *Escherichia coli* During 20,000 Generations

Richard E. Lenski,[1] Cynthia L. Winkworth,[2] Margaret A. Riley[2]

[1] Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA
[2] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA

Received: 18 July 2002 / Accepted: 4 December 2002

Abstract. We examined rates of DNA sequence evolution in 12 populations of *Escherichia coli* propagated in a glucose minimal medium for 20,000 generations. Previous work saw mutations mediated by mobile elements in these populations, but the extent of other genomic changes was not investigated. Four of the populations evolved defects in DNA repair and became mutators. Some 500 bp was sequenced in each of 36 genes for 50 clones, including 2 ancestral variants, 2 clones from each population at generation 10,000, and 2 from each at generation 20,000. Ten mutations were found in total, all point mutations including mostly synonymous substitutions and nonsynonymous polymorphisms; all 10 were found in mutator populations. We compared the observed sequence evolution to predictions based on different scenarios. The number of synonymous substitutions is lower than predicted from measured mutation rates in *E. coli,* but the number is higher than rates based on comparing *E. coli* and *Salmonella* genomes. Extrapolating to the entire genome, these data predict about 250 synonymous substitutions on average per mutator population, but only about 3 synonymous substitutions per nonmutator population, during 20,000 generations. These data illustrate the challenge of finding sequence variation among bacterial isolates that share such a recent ancestor. However, this limited variation also provides a useful baseline for research aimed at finding the beneficial substitutions in these populations.

Key words: Bacterial evolution — *Escherichia coli* — Experimental evolution — Evolutionary rate — Mutation rate — Mutator phenotype — Substitution rate

Introduction

The field of molecular evolution has led to dramatic advances in our understanding of the history of life on Earth since Zuckerkandl and Pauling (1965) proposed that macromolecular sequences evolve in a clock-like fashion, such that phylogenetic relationships can be inferred from these data. To anchor times of divergence, and to estimate underlying rates of molecular evolution, calibration of the molecular clock has often relied on paleontological data (Marshall 1990; Tavare et al. 2002).

Although historical approaches remain the major focus in evolutionary biology, recent years have witnessed considerable growth of experimental approaches, in which researchers study evolution in action in defined environments (Bennett 2002). For years, such work mostly used fruitflies (Dobzhansky and Pavlosky 1957; Rose 1984; Moya et al. 1995) but recent work has expanded to include viruses (Bull et al. 1993; Elena et al. 1998; Wichman et al. 1999; Burch and Chao 2000), bacteria (Chao and Cox 1983; Lenski et al. 1991; Travisano et al. 1995; Rainey and

Correspondence to: Richard E. Lenski; *email:* lenski@msu.edu

Travisano 1998; Velicer et al. 1998), yeast (Paquin and Adams 1982; Zeyl 2000), and even vertebrates (Reznick et al. 1997).

The use of microorganisms offers several advantages for experimental evolution, two of which are most relevant to this paper. First, experiments can last for thousands of generations, allowing significant evolutionary change and beginning to close the gap in temporal scale with some historical studies of evolution. Nonetheless, it must be realized that even thousands of generations are a mere "drop in the bucket" of evolutionary time. Second, microbial populations can be started from a genetically homogeneous clone—indeed, a single haploid cell. Thus, one can be confident that evolutionary changes reflect only genetic events that occurred during an experiment, as opposed to sorting of genetic variation that was already present in the founding population. This issue is especially relevant for interpreting cases of parallel evolution, where the level of independence between replicate lines is difficult to establish if the genetic variant in question was present in the ancestral population.

With viruses, it has become possible to obtain complete genome sequences for isolates derived from evolution experiments, and to elucidate every change from the ancestral genotype, as shown by Wichman et al. (1999). In that study, some nonsynonymous mutations were discovered to have been substituted repeatedly in replicate populations, and these parallel changes are strong candidates for beneficial mutations. Because such mutations are evidently under selection, however, one cannot directly infer the underlying mutation rate from the observed substitution rate. More generally, several studies with asexual bacteria and inbred fruitflies have begun to identify de novo mutations substituted in evolving experimental populations. Some of these studies have found beneficial mutations in candidate genes (Notley-McRobb and Ferenci 1999, 2000; Cooper et al. 2001, 2003), others measured the accumulation of certain classes of mutation to estimate their underlying rates (Nuzhdin et al. 1997; Schug et al. 1998; Maside et al. 2000), and still other studies have discovered some mutations but not yet determined their fitness effects or their underlying rates (Bergthorsson and Ochman 1999; Papadopoulos et al. 1999; Schneider et al. 2000; Riehle et al. 2001). Except for viruses, it is not yet practical to obtain whole-genome sequences for the many genotypes that are generated and must be interpreted in an evolution experiment. However, in this study, we make a start toward that goal by sequencing about 500 bp from each of 36 randomly chosen gene regions in 50 clones from 12 populations that are part of a long-term evolution experiment with the bacterium *Escherichia coli*. In all, we sequenced almost a million base pairs (36 regions × 500 bp per region × 50 clones), and our study had the power to detect substitutions in more than 4 billion bp·generations (36 regions × 500 bp per region × 12 populations × 20,000 generations per population). And because the regions were chosen at random (not candidate genes chosen for their potential role in adaptation or certain mutational classes chosen for their hypermutability), we can use data on synonymous substitutions, which are presumed to be selectively neutral, to infer typical base-pair mutation rates.

The main features of this long-term evolution experiment with *E. coli* are as follows (Lenski et al. 1991; Lenski and Travisano 1994). Twelve populations were founded, six from each of two variants that differ only in a neutral genetic marker used to distinguish competitors in assays of relative fitness. Each population was founded from a single cell. They have been propagated by serial transfer in a minimal medium supplemented with glucose at 25 μg per ml and incubated at 37°C. Every day, each population has been diluted 1:100 into fresh medium. The resulting regrowth allows about 6.6 ($=\log_2 100$) generations of binary fission per day. The ancestral strains have been kept frozen at −80°C throughout the experiment, and evolved samples are periodically saved and stored as well. In the present study, we examine 50 clones, including the 2 variants of the ancestor, 2 randomly chosen clones from each population at generation 10,000, and 2 more clones from each population at generation 20,000.

During these 20,000 generations, the average fitness of the evolved populations increased by about 70% relative to that of their ancestor, based on competition experiments performed in the same environment (Cooper and Lenski 2000). Numerous other phenotypic changes have also occurred in these evolving populations, including the production of larger cells (Lenski and Mongold 2000) and subtle, but significant, changes in catabolic functions (Cooper and Lenski 2000). Of particular relevance to this study, 3 of the 12 populations evolved defects in DNA repair by generation 10,000 (Sniegowski et al. 1997), and a fourth did so by generation 20,000 (Cooper and Lenski 2000). Owing to these defects, these "mutators" have genomewide mutation rates that are about 100-fold higher than those of the ancestor or other populations (Sniegowski et al. 1997).

Molecular genetic analyses of these populations are proceeding on various fronts. Native insertion sequence (IS) elements, which often exist in multiple copies in bacterial genomes, have been used as probes against digested genomic DNA to identify IS-mediated mutations, including insertions, deletions, and inversions (Papadopoulos et al. 1999). After 10,000 generations, each population typically possessed several IS-mediated mutations that had been substi-

tuted and a number of others that were polymorphic. In two focal populations, several of these IS-mediated substitutions were precisely characterized at the sequence level (Schneider et al. 2000); a set of parallel deletions was also characterized in all 12 populations (Cooper et al. 2001).

In this study, we seek to detect and quantify other kinds of mutations substituted in these evolving populations. While it is possible to find nearly all IS-mediated mutations in a genome, owing to the changes in DNA fragment sizes that they cause, it is not currently possible to find every point mutation except by exhaustive sequencing. Therefore, in this study, we sequenced several dozen randomly chosen gene regions in several clones from all 12 populations, to obtain some measure of the extent of genomic evolution in this experiment. We also seek to compare the observed rates of substitutions, especially those that are selectively neutral, with alternative models based on different lines of evidence (Drake 1991; Ochman et al. 1999; see also Drake et al. [1998] for a general review comparing rates of spontaneous mutation in many organisms).

Although our interest in experimental evolution has been primarily from the standpoint of addressing fundamental questions, information about short-term rates and patterns of molecular evolution has important applications in the epidemiological and forensic tracking of pathogens. Such applications are particularly difficult when potential organisms of interest share a recent common ancestor, as is the case for the *Bacillus anthracis* used in the recent bioterrorism (Read et al. 2002). Molecular studies on the long-term experimental *E. coli* populations provide data concerning the extent of genetic divergence that can evolve over a defined period in a bacterial species.

Materials and Methods

Bacterial Strains

Twelve populations of *Escherichia coli* B were founded from the same ancestor, except for a neutral Ara marker embedded in the experiment, as described below. The 12 populations were propagated by serial transfer for 20,000 generations (3000 days) in a minimal salts medium supplemented with 25 μg glucose per ml, as described previously (Lenski et al. 1991; Lenski and Travisano 1994; Cooper and Lenski 2000). Each population started from a single cell, and the populations have depended on spontaneous mutations for their subsequent evolution.

The B strain used as ancestor in this experiment was designated Bc251, and described as *T6-r str-r* rm_{111} ara^-, by S. Lederberg (1966). A spontaneous Ara^+ mutant was selected from it (Lenski et al. 1991). The Ara^- and Ara^+ variants differ by a point mutation in *araA* (C. Borland, unpublished data). Six of the 12 populations were founded by each of these two variants. The populations are designated Ara−1, Ara−2, Ara−3, Ara−4, Ara−5, Ara−6, Ara + 1, Ara + 2, Ara + 3, Ara + 4, Ara + 5, and Ara + 6. Samples from each population were periodically taken and stored at −80°C. For this study, we chose two clones at random from each population at generation 10,000 and two more at generation 20,000. We sequenced 36 gene regions (next section) in each of these 48 evolved clones as well as the 2 variant ancestors. For those samples in which one or both clones had a mutational difference from the ancestor, three additional clones were randomly chosen for further sequencing.

As will be very important for interpreting the results, the populations diverged from one another in the genetic mechanisms controlling genomewide mutation rates (Sniegowski et al. 1997; Cooper and Lenski 2000; P.D. Sniegowski, unpublished data). Eight populations retained the low ancestral mutation rate throughout the 20,000 generations. Three populations (Ara−2, Ara−4, and Ara + 3) evolved genetic defects in methyl-directed mismatch repair and they became mutators prior to generation 10,000. Another population (Ara + 6) also became a mutator, but it is enigmatic in several respects that are discussed later.

Gene Regions Sequenced

Four ORFs were randomly chosen for sequencing in each 10-min interval of the *E. coli* K-12 chromosome. Primer pairs were designed using the K-12 genome that would allow sequencing of approximately 500 bp in each gene. Four genes (*ymfE* [25.8 min], *rfbB* [45.5 min], *eutE* [55.4 min], and *yhfS* [75.5 min]) did not yield product for sequencing using the initial primers, perhaps owing to differences between the B and the K-12 genomes; these genes were not pursued further. Table 1 lists each of the 36 genes in this study, its map location, the total number of base pairs and the number sequenced, and the function of its gene product. All 36 regions (totaling 18,374 bp) were sequenced in all 50 clones (= 918,700 bp sequenced in total). Of the base pairs sequenced in each clone, 4360 (23.7%) and 14,014 (76.3%) were at risk for synonymous and nonsynonymous mutations, respectively, as calculated using the Molecular Evolutionary Analysis Package (Etsuko Moriyama, Yale University; version 6/22/2000) and based on the algorithm of Nei and Gojobori (1986).

DNA Extraction and Amplification

DNA was isolated from each clone using a Qiagen Tissue Kit, following the manufacturer's instructions, or by a boiling technique. The latter involved transferring cells from a single colony into a 1.5-ml microcentrifuge tube containing 100 μl of sterile Milli-Q water; this preparation was boiled for 10 min and stored at 4°C for subsequent manipulations. Typically 30–60 ng of nucleic acid was used for individual PCR amplifications. Oligonucleotide primers, from 18 to 25 nucleotides, were designed using the Primer3 program (www.genome.wi.mit.edu) or Mac Vector 6.5. 1 (Oxford Molecular). PCR primer pairs were designed to produce an amplified product between 500 and 600 bp in length. The primer pairs employed for all 36 genes listed in Table 1 are available on request from the authors. All reactions were performed in a Gene Amp 9700 thermocycler (PE Biosystems) and using the following program: 95°C for 5 min; 25 cycles at 95, 57, and 72°C, for 45 s each; and, finally, 72°C for 7 min. Amplified products were visualized on 1% agarose gels containing 5 ng ethidium bromide per μl to determine the correct size and concentration. The products were then purified using either a GeneClean III kit (Bio 101, Inc.) or a QIAquick 8 PCR purification kit (Qiagen) following the manufacturer's recommended procedures.

DNA Sequencing and Analysis

Sequencing reactions were prepared for analysis on an ABI 377 automated sequencer using the Big Dye Terminator Cycle Se-

Table 1. The 36 genes that were partially sequenced in this study, listed in order of their map position

Gene	Map position[a]	Total bp in ORF	No. of bp sequenced	First–last bp sequenced[b]	Function	Reference
caiC	0.8	1569	525	868–1392	Carnitine metabolism	Eichler et al. (1994)
caiT	0.9	1515	588	697–1284	Carnitine metabolism	Verheul et al. (1998)
yabH (*djlA*)	1.2	816	496	163–658	Molecular chaperone	Kelley and Georgopoulos (1997)
yabJ	1.6	699	494	151–644	Purine biosynthetic pathway	Saxild et al. (2001)
amtB	10.2	1287	495	730–1224	Ammonium uptake	Soupene et al. (1998)
ybbB	11.4	1095	495	379–873	Putative capsule anchoring protein	Sadosky et al. (1991)
yleA (*b0661*)	14.9	1425	504	820–1323	Methylthiolation modified tRNA	Esberg et al. (1999)
potI	19.3	846	504	229–732	Periplasmic transport system of putrescine	Pistocchi et al. (1993)
ycdS (*b1024*)	23.5	2424	494	919–1412	Possibly involved in haemin uptake and storage	Jones et al. (1999)
flgE	24.4	1209	500	229–728	Flagellar hook subunit protein	Komeda et al. (1978)
trpB	28.4	1194	516	607–1122	Tryptophan synthesis	Yee et al. (1996)
adhP (*b1478*)	33.4	1041	521	226–746	Alcohol dehydrogenase	Blattner et al. (1997)
ydeU (*b1509*)	34.3	1401	508	346–853	Function unknown	Blattner et al. (1997)
ydeD	34.9	801	496	121–616	Extrusion of cysteine pathway metabolites	Dassler et al. (2000)
ydhT (*b1669*)	37.7	813	510	82–591	Function unknown	Blattner et al. (1997)
hisI	45.2	612	503	61–563	PR-ATP pyrophosphatase and PR-AMP cyclohydrolase	Chiarotti et al. (1986)
yegS (*b2086*)	46.7	900	492	139–630	Function unknown	Blattner et al. (1997)
yeiU (*b2174*)	48.9	750	495	121–615	Function unknown	Blattner et al. (1997)
atoA	50.1	651	537	61–597	Acetate co-a-transferase β subunit	Rhie and Dennis (1995)
nrdA	50.5	2287	545	1573–2117	Class I ribonucleotide reductase	Gallardo-Madueno et al. (1998)
cysZ	54.5	762	518	148–665	Sulfate transport and assimilation	Britton et al. (1983)
aas	64.1	2160	486	448–933	Lysophospholipid acylation	Jackowski et al. (1994)
parC	68.2	2259	499	657–1155	Cis-acting centromere-like component of DNA segregation	Moller-Jensen et al. (2002)
yhaO	70.2	1278	524	37–560	Putative transporter or regulatory protein	Blattner et al. (1997)
folP	71.6	894	499	151–649	Dihydropteroate synthase	Enne et al. (2002)
yhfC (*gutS*)	75.2	1182	521	97–617	Function in response to selenite and tellurite	Guzzo and Dubow (2000)
feoB	76.3	2322	506	1606–2111	Ferrous iron transport protein B	Kammler et al. (1993)
kdgK	79.3	1149	474	427–900	2-Keto-3-deoxygluconate kinase	Pujic et al. (1998)
yibD	81.6	1035	499	442–940	Function unknown	Blattner et al. (1997)
polA	87.2	2787	537	1180–1716	Nonreplicative DNA polymerase I	A1 Mamun et al. (2000)
fdhE	87.9	930	504	310–813	Affects formate dehydrogenase-N	Abaibou et al. (1997)
glpX	88.6	1011	549	358–906	Involved with glycerol metabolism	Donahue et al. (2000)
hydH	90.5	1398	482	643–1124	Two-component regulatory system responsive to zinc and lead	Leonhartsberger et al. (2001)
yjcD	92.2	1350	513	631–1143	Function unknown	Blattner et al. (1997)
yjfQ	95.2	756	520	88–607	Putative transcriptional regulator	Blattner et al. (1997)
yjiN	98.4	1281	525	109–633	Function unknown	Blattner et al. (1997)

[a] Based on location in K-12 genome, in minutes (Blattner et al. 1997).
[b] Primer pairs for all 36 genes are available on request from the authors.

quencing Kit (ABI Applied Biosystems) following the manufacturer's recommended protocol. All PCR products were sequenced in both the forward and the reverse directions. Sequencing reactions were purified using a 96-well filtration block (Edge Biosystems) following the manufacturer's recommended procedure. Sequences were assembled manually and aligned using Sequencer 3.1.1 (Gene Codes). Only those regions with coverage of both strands were analyzed. In every case where a putative mutation was identified, it was confirmed or rejected by resequencing the relevant region twice using independent DNA extractions. Whenever resequencing confirmed a mutation in one or both clones, that gene region was sequenced in three more clones randomly picked from the same population and generation.

Results

We present our results separately for the eight populations that retained the low ancestral mutation rate through 20,000 generations, then for three popula-

Table 2. Three quantitative scenarios for rates of sequence evolution at synonymous sites, and their corresponding predictions for this study

	Substitution-rate scenario		
	Intermediate	Low	High
Mutation rate per bp per generation	5.40×10^{-10}	2.25×10^{-11}	5.40×10^{-9}
Expected number of synonymous substitutions			
Per nonmutator population[a]	0.0471	0.0020	0.4709
Per mutator population[b]	3.7574	0.1566	37.5739

[a] After 20,000 generations in 18,374 bp sequenced, with 23.73% of sites at risk for synonymous mutations.
[b] With a 104-fold increase in the mutation rate for 15,300 generations of 20,000 total.

tions that evolved into mutators before the midpoint of the experiment, and, finally, for one rather enigmatic population that also became a mutator. However, before presenting these data, we present three alternative mutation-rate scenarios that predict very different levels of synonymous substitutions (Table 2). We can then evaluate whether any of these scenarios are supported, or conversely excluded, by our data.

Alternative Scenarios

One of the theoretical foundations of molecular evolution is that the steady-state substitution rate of neutral mutations is independent of population size (Kimura 1983), including the influence of selection on effective population size. Moreover, the steady-state rate of substitution at neutral sites is equal to the underlying mutation rate at those sites (Kimura 1983), so that the mutation rate can be estimated simply and directly as the rate of substitution at neutral sites. In all three scenarios below, we assume that synonymous mutations are neutral. We also assume that 23.7% of all mutations are synonymous, which is the fraction corresponding to the gene regions that we sequenced.

The three scenarios differ only in their base-pair mutation rates. Each scenario uses two values: a basal value that is applied to the eight nonmutator populations and a higher value used for the three populations that evolved into imitators before the midpoint of the experiment. The latter values are 104-fold higher than the former values. We calculated this multiplicative factor as the geometric mean of eight estimates from fluctuation tests using all three of these mutator populations and three loci (Sniegowski et al. 1997). The corresponding arithmetic mean is 212-fold, but it is heavily influenced by the highest values. These three populations were mutators on average for 15,300 of the 20,000 generations (Sniegowski et al. 1997), and we apply the 104-fold higher mutation rate only to this number of generations.

Intermediate Mutation Rate. Summarizing data obtained from studies of mutation at several well-characterized loci in *E. coli*, Drake (1991) estimated an average mutation rate of 5.4×10^{-10} per bp per generation. Given the 20,000 generations elapsed and 18,374 bp sequenced, this scenario predicts 0.047 synonymous substitution in each nonmutator population (Table 2). If we sum this value over all eight nonmutator populations, we expect 0.38 synonymous substitutions. Therefore, it is somewhat more likely to find no synonymous mutations substituted than one or more substitutions.

Under this intermediate scenario, the mutator populations are predicted to average about 3.8 synonymous substitutions among the 18,374 bp sequenced (Table 2). Summing over all three mutator populations, we expect about 11.3 synonymous substitutions.

Low Mutation Rate. Ochman et al. (1999) argued that the mutation rate given by Drake (1991) is much too high, based on its inconsistency with synonymous substitution rates calculated from DNA sequence divergence between *E. coli* and *Salmonella enterica*. Ochman et al. estimated that the substitution rate per synonymous site is 4.5×10^{-9} per year. This annual rate corresponds to 2.25×10^{-11} substitutions per synonymous site per generation, assuming 200 generations per year (the middle of the range they gave for natural populations of *E. coli*). Using neutral theory, this synonymous substitution rate provides a direct estimate of the base-pair mutation rate, if one assumes that synonymous substitutions are selectively neutral.

Applying this low rate to the long-term *E. coli* populations, along with the number of base pairs sequenced and the fraction at risk for synonymous mutations, we expect to observe only about 0.002 synonymous mutation substituted per nonmutator population (Table 2). We would almost certainly not discover even a single synonymous substitution in any of the eight nonmutator populations. This low-rate scenario predicts about 0.16 synonymous substitution per mutator population (Table 2), and 0.47

such substitutions over the three mutator populations combined. Therefore, it is about as likely to find no synonymous mutations substituted in any of the mutator populations as it is to find one or more substitutions.

High Mutation Rate. At the other end of the spectrum, some colleagues have suggested to us that they would expect to see many more mutations in the long-term experiment than predicted using the mutation rate calculated by Drake (1991). Two explanations that are relevant to synonymous substitutions have been suggested. First, Drake's estimate is based on exponentially growing cells. Starving cells might have higher mutation rates than growing cells, and starving cells can accumulate mutations without the elapse of generations (Mittler and Lenski 1990; Zambrano et al. 1993; Bridges 1998). The *E. coli* populations in the long-term experiment were transferred to fresh media each day, but they depleted the available glucose in only about 8 h. Hence, the cells spent more than half their time in the stationary phase, and they may thus have experienced a higher mutation rate per generation. Second, some genes are much more mutable than others, whether by happenstance or to promote evolvability (Moxon et al. 1994). It is possible that the handful of genes used in the estimate by Drake produced a substantial underestimate of the total genomic mutation rate. For example, Cooper et al. (2001) discovered that the *rbs* operon was partially or entirely deleted in all 12 long-term *E. coli* populations. They showed that this high rate of loss was caused, in part, by hypermutability attributed to an adjacent IS*150* element. The localized mutation rate leading to this particular loss of function was much more than an order of magnitude higher than expected from the estimate of Drake (1991).

Both these explanations are quite reasonable and presumably have at least some effect. However, we are unaware of a publication that sets out such precise parameter estimates as the two preceding scenarios, although one can easily imagine that the net effect could be very large. For the sake of illustration, we increased the base-pair mutation rate by 10-fold relative to that of Drake (Table 2).

Applying this high mutation rate to our system, we would expect about 0.47 synonymous substitution per nonmutator population (Table 2), for a total of about 3.77 across all eight nonmutator populations. This scenario is the only one of the three that would be inconsistent with the failure to discover any synonymous substitutions at all in the nonmutator populations. This scenario also predicts about 38 synonymous substitutions in each mutator population (Table 2) and about 113 such substitutions in the three mutator populations combined.

Sequence Evolution

Let us now turn to the results of our sequencing to see how they compare to the three scenarios. A total of 18,374 bp, from 36 randomly chosen ORFs, was sequenced in each of 50 clones. Of the base pairs sequenced, about 23.7% were at risk for synonymous mutations. The 50 clones included the 2 ancestors that differed by a neutral marker embedded in the experimental design, 2 clones randomly sampled from each of the 12 evolving populations at generation 10,000, and 2 more clones from each population at generation 20,000. The two ancestral variants were identical at every base, as the genetic marker lies in the *araA* gene that was not sequenced in this study; therefore, we subsequently refer simply to the ancestor.

In every case where a sequence difference was initially observed between a derived clone and the ancestor, that difference was further tested by resequencing the relevant gene at least twice. Also, whenever resequencing confirmed a mutation in one or both clones in the same sample, that gene region was sequenced in three additional clones randomly picked from the same population and generation. We define a *substitution* operationally as all five clones in a sample sharing the same mutational difference from the ancestor. In contrast, we define a *polymorphism* as any case in which between one and four of the five clones in a sample share a particular mutation.

Nonmutator Populations. Eight of the 12 populations retained ancestral rates throughout the 20,000 generations of experimental evolution (Sniegowski et al. 1997; Cooper and Lenski 2000; P.D. Sniegowski, unpublished data).

Among the 36 randomly chosen gene regions that we sequenced, not a single substitution was observed in any of these eight populations, nor were any polymorphisms found. These data are consistent with both the intermediate and the low mutation-rate scenarios. However, they are inconsistent with the high mutation-rate scenario ($p = 0.0231$ based on the Poisson expectation for the zero-class).

Mutator Populations. Three of the 12 populations evolved defects in their methyl-directed mismatch repair pathway by 10,000 generations, and they remained defective throughout the experiment (Sniegowski et al. 1997; Cooper and Lenski 2000; P.D. Sniegowski, unpublished data). On average, these populations were mutators for 15,300 of the 20,000 generations, and during this time their mutation rates were elevated by about 100-fold relative to that of the ancestor.

Table 3. Mutations found in 12 evolving populations, among the 36 genes partially sequenced in this study

Gene	Nucleotide position[a]	Mutation	Amino acid change	Population	Generation	Frequency (%)[b]
yjcD	264	A → G	None	Ara−4	20,000	100
adhP	400	T → C	None	Ara−4	20,000	100
ybbB	454	T → C	Tyr → His	Ara−4	20,000	100
nrdA	432	G → A	None	Ara+3	20,000	100
yjiN	383	A → C	Asp → Ala	Ara+6	10,000	20
feoB	434	T → G	Val → Gly	Ara+6	10,000	40
feoB	278	T → G	Phe → Cys	Ara+6	20,000	20
caiT	341	A → C	Glu → Ala	Ara+6	20,000	80
ycdS	220	A → C	Asn → His	Ara+6	20,000	20
hisI	169	A → C	Ile → Leu	Ara+6	10,000	100
hisI	169	A → C	Ile → Leu	Ara+6	20,000	80

[a] Within the K-12 ORF.
[b] Based on five clones chosen at random.

Among the 36 random gene regions sequenced, we found a total of four substitutions in these populations, where a substitution indicates that all five clones of five in a population sample shared the same mutational difference from the ancestor (Table 3). Population Ara−4 had three of the substitutions, including a synonymous A → G mutation in *yjcD*, a synonymous T → C mutation in *adhP*, and a nonsynonymous T → C mutation in *ybbB*. Ara + 3 had a synonymous G → A mutation in *nrdA*. No substitutions were found in the other mutator population, Ara−2. All four of these mutations are either A·T → G·C or G·C → A·T transitions, which are precisely the classes of point mutation that are known to be overrepresented as a consequence of defects in methyl-directed mismatch repair (Friedberg et al. 1995, p. 376).

Overall, three of the four substitutions were synonymous, and all four were substituted late, with none of them detected even as a polymorphism after 10,000 generations. In fact, no polymorphisms were seen in the 36 random genes in any of the three mutator populations at either time point.

The data for the three mutator populations fall in between the expectations based on the low and intermediate mutation-rate scenarios, and they are thoroughly inconsistent with the high scenario (Table 2). Based on the Poisson distribution, the probability of observing three or fewer synonymous substitutions under the intermediate scenario is 0.0040, meaning that significantly fewer substitutions were observed than would be consistent with this scenario. The probability of observing three or more such substitutions under the low scenario is 0.0122, indicating that there were significantly more such substitutions than would be compatible with that scenario. Recall, however, that we used the lower geometric-mean estimate (104-fold) of the effect of the mutator status on mutation rate. If we use instead the higher arithmetic-mean estimate (212-fold), then the low mutation-rate scenario predicts a total of 0.96 synonymous substitution in the mutator populations and it is marginally consistent with the observed number ($p = 0.0726$ based on the Poisson distribution).

An Enigmatic Population. Population Ara + 6 exhibited some unique features before we began this study, and it has become even more enigmatic based on our sequencing. Ara + 6 was not identified as a mutator at generation 10,000 (Sniegowski et al. 1997), but it had become a mutator by generation 20,000 (Cooper and Lenski 2000; P.D. Sniegowski, unpublished data). Unlike the three other mutators described above, the defect in Ara + 6 was not genetically complemented by wild-type alleles of any gene in the methyl-directed mismatch repair pathway (P.D. Sniegowski, unpublished data). Ara + 6 is also unusual in that it produces sickly colonies (small size and low plating efficiency) on our standard rich plating medium, although it is highly competitive in the liquid minimal-glucose medium in which it evolved (as are all 12 evolved populations). This sickly colony appearance was manifest before 10,000 generations, whereas the mutator phenotype seems not to have arisen until much later.

Among the 36 gene regions sequenced, we found six different mutations in Ara + 6, which is more than in all the other populations combined (Table 3). The mutations in Ara + 6 include a nonsynonymous A → C mutation in *yjiN,* two different nonsynonymous T → G mutations in *feoB*, a nonsynonymous A → C mutation in *caiT*, a nonsynonymous A → C mutation in *ycdS*, and a nonsynonymous A → C mutation in *hisI*.

The pattern of variation in Ara + 6 is also unlike that in the other populations in several respects. (1) Three of the six mutations in A + 6 were present in the 10,000-generation sample, whereas all four other mutations were found only at generation 20,000. (2) All six mutations in Ara + 6 were nonsynonymous, whereas three of four in other populations were synonymous. This difference is significant

($p = 0.0333$, two-tailed Fisher's exact test). (3) None of the six mutations in Ara + 6 led to a substitution, whereas all four mutations found in the other populations were substituted. The situation is complicated with *hisI* in Ara + 6 because all five clones from generation 10,000 had the mutation, but only four of five clones at generation 20,000 carried it. Even if we count the *hisI* mutation as having been substituted, the difference between genetic polymorphisms in Ara + 6 and substitutions in the other populations is significant ($p = 0.0476$, two-tailed Fisher's exact test). (4) All six mutations found in Ara + 6 are A·T → C·G transversions, while all four mutations in other populations are transitions. This difference is highly significant ($p = 0.0048$, two-tailed Fisher's exact test).

Based on the abundance, timing, and other patterns of mutation in Ara + 6, we hypothesize that this population became a mutator before generation 10,000 and that this phenotype was the consequence of a lesion in the *mutT* gene. The failure to discern its mutator status at that time could mean that Ara + 6 was polymorphic for the mutator phenotype or, perhaps, the population reverted between mutator and nonmutator phenotypes during its history. In any case, mutators caused by defects in *mutT* promote A·T → C·G transversion mutations (Friedberg et al. 1995, p. 159), precisely those seen in Ara + 6. Moreover, *mutT* mutators have more potent mutagenic effects than those caused by defects in methyl-directed mismatch repair (Friedberg et al. 1995). This potency would explain the higher number of mutations in Ara + 6 than in the other mutator populations. It could also account for the pattern in Ara + 6 of polymorphic and nonsynonymous mutations, which are expected in a population that suffers from an elevated load of deleterious mutations, although one would also expect to observe more synonymous substitutions as well. We suggest that future work be directed toward testing the hypothesis that Ara + 6 has been a *mutT* mutator during part of its history.

Discussion

We examined the extent of DNA sequence variation in 12 populations of *E. coli* that evolved for 20,000 generations in a simple laboratory environment. We sequenced 36 gene regions, about 500 bp in length, in each of 48 evolved clones and their ancestor. Only 10 mutations were found in total, and all 10 were in four populations that had been mutators for much of the experiment owing to evolved defects in DNA repair pathways. Four of the 10 mutations were substitutions in populations that had evolved defects in methyl-directed mismatch repair, and these mutations were A·T → G·C or G·C → A·T transitions, which are characteristic of this class of mutators. The other six mutations were polymorphic in one population that evolved an uncharacterized mutator phenotype. All six of these mutations were A·T → C·G transversions, which strongly suggests that this population evolved a defect in *mutT*. No mutations at all were found in any of the eight populations that kept the low ancestral mutation rate throughout the 20,000 generations, among the 18,374 bp sequenced for each population.

We used these data to evaluate three scenarios for rates of molecular evolution, focusing on synonymous substitutions that we assume are selectively neutral. Using the neutral theory of molecular evolution (Kimura 1983), the base-pair mutation rate and the proportion of mutations that are neutral are sufficient to predict the steady-state rate of neutral substitutions, regardless of population size and the effect of selection thereon. Our intermediate scenario was based on the widely cited estimate of the *E. coli* mutation rate given by Drake (1991). The low mutation-rate scenario followed from Ochman et al. (1999), who reasoned from comparative data that the mutation rate given by Drake must be too high. Our high scenario increased the mutation rate 10-fold relative to that of Drake, a difference that could reflect mutations during stationary phase or scattered hypermutable loci.

The high mutation-rate scenario is decisively rejected by the sequence data, because it predicted many more synonymous substitutions than were observed. The intermediate scenario also predicted significantly more synonymous substitutions than were seen among the mutator populations, whereas the low scenario predicted fewer synonymous substitutions than observed in those populations. An even better scenario, within the context of this particular experiment, would have a base-pair mutation rate between the value of Drake (1991) and that of Ochman et al. (1999). In particular, our best estimate of the ancestral mutation rate in the repair-proficient populations is 1.44×10^{-10} per bp per generation. This estimate is back-calculated from the substitution rate in the mismatch-repair defective mutator populations, using a 104-fold difference in mutation rates.

We can also use the substitutions in randomly chosen genes to extrapolate to the genome as a whole. Given that we observed a total of three synonymous substitutions among 18,374 bp sequenced in the three mismatch-repair mutator populations, and using a total genome size of 4.64×10^6 bp (Blattner et al. 1997), this extrapolation implies an average of 253 genomic synonymous substitutions per mutator population after 20,000 generations. These populations experienced a 104-fold higher mutation rate for 15,300 generations, on average, than did the eight populations that retained the low ancestral mutation rate throughout. Taking this difference into account,

the average nonmutator population is predicted to have had only 3.2 synonymous substitutions in its entire genome! (If we had used the arithmetic-mean estimate for the mutator effect instead of the geometric-mean, the predicted number of substitutions in the nonmutator populations would be even lower.) These genomic extrapolations are subject to uncertainty owing to the small number of synonymous substitutions discovered. In particular, the 95% confidence interval for a count of 3 extends from 0.62 to 8.78, and the confidence intervals for the numbers of substitutions are proportional. If we apply this uncertainty to the nonmutator populations to obtain an upper bound, the average population has accumulated fewer than 10 synonymous substitutions in its genome during 20,000 generations.

We emphasize that these numbers reflect synonymous substitutions only. Three of the four substitutions observed in the randomly chosen genes are synonymous mutations; the fourth is a nonsynonymous mutation in *ybbB,* which encodes a putative capsule anchoring protein (Sadosky et al. 1991) that might be under relaxed selection in the laboratory. Three lines of evidence demonstrate that there have been a number of other substitutions—including ones that are beneficial and nonsynonymous—in all the populations, including nonmutator as well as mutator.

First, the populations have improved their fitness relative to the ancestor by about 70%, on average, during the 20,000 generations (Cooper and Lenski 2000). Based on the step-like dynamics of adaptation during the first 2,000 generations, no substitution accounted for more than about a 10% improvement, and the more gradual later gains imply more mutations of smaller effect (Lenski et al. 1991; Lenski and Travisano 1994; Gerrish and Lenski 1998). Thus, there must be at least 10, and perhaps 20 or so, beneficial substitutions in each population. It is not surprising, however, that sequencing less than 0.5% of the genome would have missed these beneficial substitutions.

Second, previous work with these populations found numerous IS-mediated mutations, including insertions, deletions, and inversions (Papadopoulos et al. 1999; Schneider et al. 2000; Cooper et al. 2001). One particular set of IS-mediated mutations, involving deletion of part or all of the *rbs* operon, was substituted in all 12 of the populations, although the physical extent of the deletion varied in each case (Cooper et al. 2001). Again, however, it is not surprising that 10 or 20 IS-mediated substitutions per population would be missed by sequencing less than 0.5% of the genome.

Third, members of the REL laboratory have found four genes to date in which at least 8 of the 12 independently evolved populations have nonsynonymous substitutions. Two of these cases, involving deletions of the *rbs* operon (Cooper et al. 2001) and point mutations in the *spoT* gene (Cooper et al. 2003), have been published so far; and in both studies the substitutions were shown to be beneficial on the ancestral background by constructing otherwise isogenic strains. The other two cases are presently being studied and will be reported in future papers. These four cases indicate that the beneficial mutations are concentrated in a relatively few genes, which makes it unlikely that they would be discovered in a study such as the present one. However, the present study provides a valuable statistical control, or baseline, for the cases where multiple populations show similar nonsynonymous substitutions.

Finally, in closing, we would briefly mention the relevance of our study for molecular epidemiology, including the ongoing investigation of the source of the *Bacillus anthracis* used in recent bioterrorism (Read et al. 2002). One of the challenges that can arise in such work is when two potential sources are very closely related by a recent common ancestor. It is rather daunting to realize that after 20,000 bacterial generations probably fewer than 10 synonymous mutations were substituted in the entire genome in an evolving *E. coli* population that retained its DNA repair functions.

Acknowledgments. We thank Derek Smith and Carla Valletta-Goldstone for excellent technical assistance; Paul Sniegowski and Chris Borland for generously sharing unpublished data; Tim Cooper, Santiago Elena, and Howard Ochman for valuable discussions; and two anonymous reviewers for comments. This study was supported by a grant from the U.S. National Science Foundation to R.E.L. and M.A.R.

References

Abaibou H, Giordano G, Mandrand-Berthelot MA (1997) Suppression of *Escherichia coli* formate hydrogenlyase activity by trimethylamine N-oxide is due to drainage of the inducer formate. Microbiology 143:2657–2664

Al Mamun AA, Yadava RS, Ren L, Humayun MZ (2000) The *Escherichia coli* UVM response is accompanied by an SOS-independent error-prone DNA replication activity demonstrable in vitro. Mol Microbiol 38:368–380

Bennett AF (2002) Experimental evolution. In: Pagel M, Frank S, Godfray C, Hall BK, Hawkes K, Hillis DM, Kodric-Brown A, Lenski RE, Pomiankowski A (eds) Encyclopedia of evolution. Oxford University Press, New York, pp 339–342

Bergthorsson U, Ochman H (1999) Chromosomal changes during experimental evolution in laboratory populations of *Escherichia coli.* J Bacteriol 181:1360–1363

Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K12. Science 277:1453–1474

Bridges BA (1998) The role of DNA damage in stationary phase ('adaptive') mutation. Mutat Res 408:1–9

Britton P, Boronat A, Hartly DA, Jones-Mortimer MC, Kornberg HL, Parra F (1983) Phosphotransferase-mediated regulation of carbohydrate utilization in *Escherichia coli* K-12: Location of the *gsr* (*tgs*) and *iex* (*crr*) genes by specialized transduction. J Gen Microbiol 129:349–356

Bull JJ, Cunningham CW, Molineux IJ, Badgett MR, Hillis DM (1993) Experimental molecular evolution of bacteriophage T7. Evolution 47:993–1007

Burch CL, Chao L (2000) Evolvability of an RNA virus is determined by its mutational neighbourhood. Nature 406:625–628

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

Chiarotti L, Alifano P, Carlomango MS, Bruni CB (1986) Nucleotide sequence of the *Escherichia coli hisD* gene and of the *Escherichia coli* and *Salmonella tvphimurium hisIE* region. Mol Gen Genet 203:382–388

Cooper TF, Rozen DE, Lenski RE (2003) Parallel changes in gene expression after 20,000 generations of evolution of *E. coli*. Proc Natl Acad Sci USA 100:1072–1077

Cooper VS, Lenski, RE (2000) The population genetics of ecological specialization in evolving *E. coli* populations. Nature 407:736–739

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

Dassler T, Maier T, Winterhalter C, Böck A (2000) Identification of a major facilitator protein *Escherichia coli* involved in efflux of metabolites of the cysteine pathway. Mol Microbiol 36:1101–1112

Dobzhansky T, Pavlosky O (1957) An experimental study of interaction between genetic drift and natural selection. Evolution 11:311–319

Donahue JL, Bownas JL, Niehaus WG, Larson TJ (2000) Purification and characterization of *glpX*-encoded fructose-1,6-bisphosphatase, a new enzyme of the glycerol-3-phosphate regulon of *Escherichia coli*. J Bacteriol 182:5624–5627

Drake JW (1991) A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci USA 88:7160–7164

Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148:1667–1686

Eichler K, Bourgis F, Buchet A, Klebler HP, Mandrand-Berthelot MA (1994) Molecular characterization of the *cai* operon necessary for carnitine metabolism in *Escherichia coli*. Mol Microbiol 13:775–786

Elena SF, Davila M, Novella IS, Holland JJ, Domingo E, Moya A (1998) Evolutionary dynamics of fitness recovery from the debilitating effects of Muller's ratchet. Evolution 52:309–314

Enne VI, King A, Livermore DM, Hall LM (2002) Sulfonamide resistance in *Haemophilus influenzae* mediated by acquisition of *sul2* or a short insertion in chromosomal *folP*. Antimicrob Agents Chemother 46:1934–1939

Esberg B, Leung HCE, Tsui HCT, Bjoerk GR, Winkler ME (1999) Identification of the *miaB* gene involved in methylthiolation of isopentenylated A37 derivatives in the tRNA of *Salmonella typhimurium* and *Escherichia coli*. J Bacteriol 181:7256–7265

Friedberg EC, Walker GC, Siede W (1995) DNA repair and mutagenesis. ASM Press, Washington, DC

Gallardo-Madueno R, Leal JF, Dorado G, Holmgren A, Lopez-Barea J, Pueyo C (1998) In vivo transcription of *nrdAB* operon and of *grxA* and *fpg* genes is triggered in *Escherichia coli* lacking both thioredoxin and glutaredoxin-1 or thioredoxin and glutathione, respectively. J Biol Chem 273:18382–18388

Gerrish PJ, Lenski RE (1998) The fate of competing beneficial mutations in an asexual population. Genetica 102/103:127–144

Guzzo J, Dubow MS (2000) A novel selenite- and tellurite-inducible gene in *Escherichia coli*. Appl Environ Microbiol 66:4972–4978

Jackowski S, Jackson PD, Rock CO (1994) Sequence and function of the *aas* gene in *Escherichia coli*. J Biol Chem 269:2921–2928

Jones HA, Lillard JW, Perry RD (1999) HmsT, a protein essential for expression of the haemin storage (Hms^+) phenotype of *Yersinia pestis*. Microbiology 145:2117–2128

Kammler M, Schon C, Hantke K (1993) Characterization of the ferrous iron uptake system of *Escherichia coli*. J Bacteriol 175:6212–6219

Kelley WL, Georgopoulos C (1997) Positive control of the two-component RcsC/B signal transduction network by DjlA: A member of the DnaJ family of molecular chaperones in *Escherichia coli*. Mol Microbiol 25:913–931

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

Komeda Y, Silverman M, Matsumura P, Simon MI (1978) Genes for the hook-basal body proteins of the flagellar apparatus in *Escherichia coli*. J Bacteriol 134:655–667

Lederberg S (1966) Genetics of host-controlled restriction and modification of deoxyribonucleic acid in *Escherichia coli*. J Bacteriol 91:1029–1036

Lenski RE, Mongold JA (2000) Cell size, shape, and fitness in evolving populations of bacteria. In: Brown JH, West GB (eds) Scaling in biology. Oxford University Press, Oxford, pp 221–235

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

Lenski RE, Rose MR, Simpson SC, Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am Nat 138:1315–1341

Leonhartsberger S, Huber A, Lottspeich F, Bock A (2001) The *hydH/G* genes from *Escherichia coli* code for a zinc and lead responsive two-component regulatory system. J Mol Biol 307:93–105

Marshall CR (1990) The fossil record and estimating divergence times between lineages: Maximum divergence times and the importance of reliable phylogenies. J Mol Evol 30:400–408

Maside X, Assimacopoulos S, Charlesworth B (2000) Rates of movement of transposable elements on the second chromosome of *Drosophila melanogaster*. Genet Res 75:275–284

Mittler JE, Lenski RE (1990) New data on excisions of Mu from *E. coli* MCS2 cast doubt on directed mutation hypothesis. Nature 344:173–175

Moller-Jensen J, Jensen RB, Lowe J, Gerdes K (2002) Prokaryotic DNA segregation by an actin-like filament. EMBO J 21:3119–3127

Moxon ER, Rainey PB, Nowak MA, Lenski RE (1994) Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr Biol 4:24–33

Moya A, Galiana A, Ayala FJ (1995) Founder-flush speciation theory: Failure of experimental corroboration. Proc Natl Acad Sci USA 92:3983–3986

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. Mol Biol Evol 3:418–426

Notley-McRobb L, Ferenci T (1999) Adaptive *mgl*-regulatory mutations and genetic diversity evolving in glucose-limited *Escherichia coli* populations. Environ Microbiol 1:33–43

Notley-McRobb L, Ferenci T (2000) Experimental analysis of molecular events during mutational periodic selections in bacterial evolution. Genetics 156:1493–1501

Nuzhdin SV, Pasyukova EG, Mackay TFC (1997) Accumulation of transposable elements in laboratory lines of *Drosophila melanogaster*. Genetica 100:167–175

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

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

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

Pistocchi R, Kashiwagi K, Miyamoto S, Sadakata Y, Kobayashi H, Igarashi K (1993) Characteristics of the operon for a putrescine transport system that maps at 19 minutes on the *Escherichia coli* chromosome. J Biol Chem 268:146–152

Pujic P, Dervyn R, Sorokin A, Ehrich SD (1998) The *kdgRKAT* operon of *Bacillus subtilis:* Detection of the transcript and regulation by *kdgR* and *ccpA* genes. Microbiology 144:3111–3118

Rainey PB, Travisano M (1998) Adaptive radiation in a heterogeneous environment. Nature 394:69–72

Read TD, Salzberg SL, Pop M, Shumway M, Umayam L, Jiang L, Holtzapple E, Busch JD, Smith KL, Schupp JM, Solomon D, Keim P, Fraser CM (2002) Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. Science 296:2028–2033

Reznick DN, Shaw FH, Rodd FH, Shaw RG (1997) Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). Science 275:1934–1937

Rhie HG, Dennis D (1995) Role of *fadR* and *atoC* (Con) mutations in poly(3-hydroxybutyrate-co-3-hydroxyvalerate) synthesis in recombinant *pha*$^{+}$ *Escherichia coli*. Appl Environ Microbiol 61:2487–2492

Riehle MM, Bennett AF, Long AD (2001) Genetic architecture of thermal adaptation in *Escherichia coli*. Proc Natl Acad Sci USA 98:525–530

Rose MR (1984) Laboratory evolution of postponed senescence in Drosophila melanogaster. Evolution 38:1004–1010

Sadosky AB, Gray JA, Hill CW (1991) The RhsD-E subfamily of *Escherichia coli* K-12. Nucleic Acids Res 19:7177–7183

Saxild HH, Brunstedt K, Neilsen KI, Jarmer H, Nygaard P (2001) Definition of the *Bacillus subtilis* PurR operator using genetic and bioinformatic tools and expansion of the PurR regulon with *glyA, guaC, pbuG, xpt-pbuX, yqhZ-folD,* and *pbuO*. J Bacteriol 183:6175–6183

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

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF (1998) The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. Mol Biol Evol 15:1751–1760

Sniegowski PD, Gerrish PJ, Lenski RE (1997) Evolution of high mutation rates in experimental populations of *Escherichia coli*. Nature 387:703–705

Soupene E, He L, Yan D, Kustu S (1998) Ammonia acquisition in enteric bacteria: physiological role of the ammonium/methylammonium transport B (*amtB*) protein. Proc Natl Acad Sci USA 95:7030–7034

Tavare S, Marshall CR, Will O, Soligo C, Martin RD (2002) Using the fossil record to estimate the age of the last common ancestor of extant primates. Nature 416:726–729

Travisano M, Mongold JA, Bennett AF, Lenski RE (1995) Experimental tests of the roles of adaptation, chance, and history in evolution. Science 267:87–90

Velicer GJ, Kroos L, Lenski RE (1998) Loss of social behaviors by *Myxococcus xanthus* during evolution in an unstructured habitat. Proc Natl Acad Sci USA 95:12376–12380

Verheul A, Wouters JA, Rombouts FM, Abee T (1998) A possible role of ProP, ProU and CaiT in osmoprotection of *Escherichia coli* by carnitine. J Appl Microbiol 85:1036–1046

Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ (1999) Different trajectories of parallel evolution during viral adaptation. Science 285:422–424

Yee MC, Horn V, Yanofsky C (1996) On the role of helix 0 of the tryptophan synthetase alpha chain of *Escherichia coli*. J Biol Chem 271:14754–14763

Zambrano MM, Siegele DA, Almiron M, Tormo A, Kolter R (1993) Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. Science 259:1757–1760

Zeyl C (2000) Budding yeast as a model organism for population genetics. Yeast 16:773–784

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) Evolving genes and proteins. Academic Press, New York, pp 97–166