Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*

Tim F. Cooper*, Daniel E. Rozen†, and Richard E. Lenski

Center for Microbial Ecology, Michigan State University, East Lansing, MI 48824

Edited by Allan Campbell, Stanford University, Stanford, CA, and approved December 5, 2002 (received for review July 22, 2002)

Twelve populations of *Escherichia coli*, derived from a common ancestor, evolved in a glucose-limited medium for 20,000 generations. Here we use DNA expression arrays to examine whether gene-expression profiles in two populations evolved in parallel, which would indicate adaptation, and to gain insight into the mechanisms underlying their adaptation. We compared the expression profile of the ancestor to that of clones sampled from both populations after 20,000 generations. The expression of 59 genes had changed significantly in both populations. Remarkably, all 59 were changed in the same direction relative to the ancestor. Many of these genes were members of the cAMP-cAMP receptor protein (CRP) and guanosine tetraphosphate (ppGpp) regulons. Sequencing of several genes controlling the effectors of these regulons found a nonsynonymous mutation in *spoT* in one population. Moving this mutation into the ancestral background showed that it increased fitness and produced many of the expression changes manifest after 20,000 generations. The same mutation had no effect on fitness when introduced into the other evolved population, indicating that a mutation of similar effect was present already. Our study demonstrates the utility of expression arrays for addressing evolutionary issues including the quantitative measurement of parallel evolution in independent lineages and the identification of beneficial mutations.

Parallel and convergent changes across lineages are hallmarks of adaptive evolution (1–7). In any single population it is difficult to determine whether a particular evolved change resulted from natural selection, but the repeated finding of the same or similar change in multiple lineages is a strong indicator that it is adaptive (4, 8). Such analyses have typically used morphological traits to assess the similarity of organisms (9–11). However, this approach is often limited by the number of traits that can be measured, making it difficult to evaluate the extent of concordant changes. Also, it is very difficult to relate most morphological changes to the underlying genetic changes.

DNA macroarrays allow one to measure gene expression simultaneously for all genes in the genome of an organism (5, 12–15). These measurements yield an expression profile of an organism, a composite phenotype with which parallelism can be assessed over thousands of "expression" traits. Moreover, because of the biochemical proximity of the expression phenotype of an organism to its genotype, parallel changes in these traits can be used to identify candidate genes that may have contributed to adaptation (16, 17).

Twelve populations of *Escherichia coli* were founded from two ancestral variants and propagated for 20,000 generations in a glucose-limited environment (18). During this time, the populations evolved and adapted to their environment, exhibiting substantial gains in competitive fitness relative to their ancestor when measured under the same culture conditions (19, 20). The fitness trajectories of the 12 populations were similar, but it is unclear to what extent this parallelism extends to the underlying physiology and genetics (21). In this study, we address this issue using macroarrays to evaluate the extent of parallel expression changes in two of these populations.

Methods

Strains and Culture Conditions. Twelve populations of *E. coli* B were founded from two genotypes that differed only by a neutral marker and then propagated for 20,000 cell generations at 37°C in a glucose-limited minimal medium (18–20). Expression arrays were obtained for five genotypes including clones isolated from two populations (designated Ara−1 and Ara+1) after 20,000 generations; their two respective ancestral clones (Ara^- and Ara^+); and an ancestral clone into which we moved a *spoT* mutation from Ara−1 (Ara^- *spoT*). Clones from the 10 other 20,000-generation populations were also used for sequencing *spoT*. Neither population used in the expression arrays was a mutator, whereas four other populations acquired mutations that substantially increased their genome-wide mutation rates (20, 22).

RNA Isolation and Macroarrays. Transcriptional profiles were obtained by using Panorama *E. coli* cDNA macroarray membranes (Sigma–Genosys). Clones were acclimated to the same culture conditions as in the evolution experiment and then diluted 1:100 into fresh medium and grown to midexponential phase before RNA extraction. This phase was chosen because previous work has shown that exponential growth rate is the most important demographic target of selection in these populations (23). Cells were harvested by using 0.45-μm filtration units (Nalge) and then resuspended in a 1:1 mix of buffer and RNAlater stabilizing solution (Ambion, Austin, TX). RNA was extracted with the RNeasy system (Qiagen, Valencia, CA) following manufacturer instructions. DNase treatment was performed by using the Qiagen on-column kit. Subsequent cDNA production, labeling, and hybridization followed the instructions of the array-membrane manufacturer. Membranes were exposed to Kodak phosphorimager screens for 24 h and then scanned on a STORM 840 PhosphorImager (Molecular Dynamics). Image files were processed by using ARRAYVISION 6.0 software (Imaging Research, St. Catherine's, ON, Canada), and the output was exported to Microsoft EXCEL 2000 for manipulation. The data sets used in this article are available at http://myxo.css.msu.edu/ecoli/arrays.

Statistical Analyses. For each array, we measured the average background count. Each gene is represented in duplicate on an array, and we subtracted the average background from the mean of the two readings to calculate the adjusted expression level for every gene. (Lowly expressed genes occasionally gave readings slightly below the background on a given array; however, no gene gave values consistently below background.) For each array we summed the adjusted expression levels over all 4,290 genes and then divided the adjusted expression for each gene by that total to obtain its standardized expression. Finally, standardized

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: ppGpp, guanosine tetraphosphate; CRP, cAMP receptor protein.

*To whom correspondence should be addressed. E-mail: cooperti@msu.edu.

†Present address: Department of Genetics, Wageningen University, 6703 BD Wageningen, The Netherlands.

www.pnas.org/cgi/doi/10.1073/pnas.0334340100

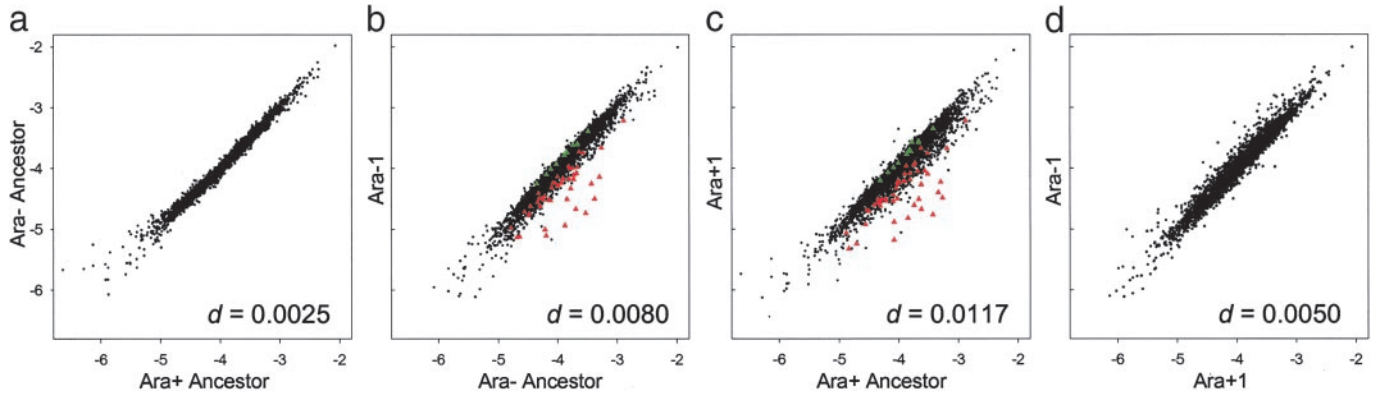

Fig. 1. Pairwise comparisons of gene expression between evolved and ancestral clones. Both axes are $\log_{10}$-transformed standardized expression levels. (*a*) Comparison of Ara$^-$ and Ara$^+$ ancestors. (*b*) Comparison of the Ara$-$1 clone from 20,000 generations and its ancestor. (*c*) Comparison of the Ara$+$1 clone from 20,000 generations and its ancestor. (*d*) Comparison of evolved Ara$-$1 and Ara$+$1 clones. (*b* and *c*) The colored points mark genes with expression that changed significantly ($P < 0.05$) relative to the ancestor in both evolved clones; genes with increased expression are shown in green, and genes with reduced expression are shown in red. The *d* values measure the overall divergence in expression profiles and were calculated as explained in *Methods*.

values were $\log_{10}$-transformed, and *t* tests were performed by using the transformed values from four independently replicated arrays per genotype to compare expression levels for two genotypes.

The divergence in expression, *d*, between genotypic pairs was calculated as $1 - r$, in other words as the deviation from perfect correlation. Each calculation of *d* used only those genes that had reproducible expression levels (coefficient of variation <0.2) in both genotypic profiles. It is important to emphasize that exclusion was based on the lack of repeatability of measurements for a gene in a particular background regardless of the influence of the gene on the correlation between genotypes. More than 1,300 genes satisfied this criterion for each pairwise comparison. The 95% confidence intervals for *d* values were calculated by using the jackknife method (24, 25) with whole arrays serving as the unit of replication; statistical comparisons of *d* values also used the corresponding pseudovalues.

DNA Sequencing. Primers were designed to allow amplification of overlapping fragments that covered each of the *crp*, *cyaA*, *gppA*, *ptsG*, *relA*, *rpoB*, and *spoT* genes in their entirety including their known upstream regulatory elements. Purified PCR products were sequenced by using an Applied Biosystems Prism 3100 automated sequencer. All mutations were confirmed by sequencing both strands at least twice.

Strain Construction. The *spoT* mutation found in Ara$-$1 was moved to the Ara$^-$ ancestral background by using a suicide plasmid-mediated approach as described (26). Briefly, a PCR product containing the mutation was cloned into pDS132 (a derivative of pCVD442, D. Schneider, personal communication). This plasmid was introduced into the Ara$^-$ ancestor, and chloramphenicol-resistant cells (formed by integration of the nonreplicative plasmid into the chromosome) were selected. Resistant clones were then streaked on LB + sucrose agar to select for cells that lost the plasmid; the plasmid expresses the *sacB* gene, making the host susceptible to killing by sucrose. These plasmid-free cells then were screened for the presence of the mutant *spoT* allele by a PCR/restriction fragment length polymorphism approach using the enzyme *Hin*4I (Fermentas, Vilnius, Lithuania) to distinguish between ancestral and evolved alleles. Putative allelic replacements were confirmed by sequencing the *spoT* gene. To control for possible mutations that might occur elsewhere in the chromosome during strain construction, we also replaced the introduced *spoT* mutation with the ancestral allele (thus undoing the mutant allelic replacement) and compared the fitness of this reconstructed ancestor to the true ancestor. We used a mutant *spoT* replacement strain for which this secondary replacement restored competitive fitness to its original level; some other replacements had lower fitness than the original state, indicating deleterious secondary mutations. These forward and reverse replacements were done several times to confirm the effect of the *spoT* mutation on fitness. The same approach was followed to move the mutant *spoT* allele from Ara$-$1 into a 2,000-generation clone of the Ara$+$1 population and to confirm the absence of secondary mutations by replacing the ancestral allele.

Fitness Assays. Following the same protocol used to measure fitness gains in the evolved populations (18–20), we performed six competitions between Ara$^-$ *spoT* and Ara$^+$ and six more between Ara$^-$ and Ara$^+$ to measure the fitness effect of the *spoT* mutation. The Ara marker is itself selectively neutral, but it allows us to distinguish competitors. The effect of the *spoT* mutation is tested by comparing the two sets of competitions. We also ran competitions using the 2,000-generation clone of Ara$+$1 with the wild-type and mutant *spoT* alleles, each with 6-fold replication, against a 2,000-generation clone of Ara$-$1. Again, the effect of the *spoT* allele is tested by comparing the two sets of competitions. In all cases, fitness is calculated as the ratio of the net population growth rates of two clones during competition (18–20).

Results and Discussion

Extent of Expression Profile Change over Time. We measured the global expression profile of clones from two independently evolved populations after 20,000 generations and their respective ancestors. For each of these four genotypes, we ran four arrays (using independent cell cultures and RNA isolations) to provide statistical replication. Fig. 1 summarizes the overall pattern of evolutionary changes in gene expression. Each panel shows the mean of four $\log_{10}$-transformed, standardized expression levels for all 4,290 genes in a pair of genotypes. Fig. 1*a* shows the very similar expression profiles of the two ancestral clones, which differ only by a single genetic marker that is selectively neutral under the conditions used (18). Their divergence in expression, *d*, is only 0.0025; this value represents the baseline noise from experimental variation. Fig. 1 *b* and *c* each show the expression patterns for one of the evolved clones and its corresponding ancestor. Note that the divergence in gene expression is much larger in both these cases ($d = 0.0117$ and 0.0080 for Ara$+$1 and Ara$-$1, respectively) than in the ancestral comparison. Both

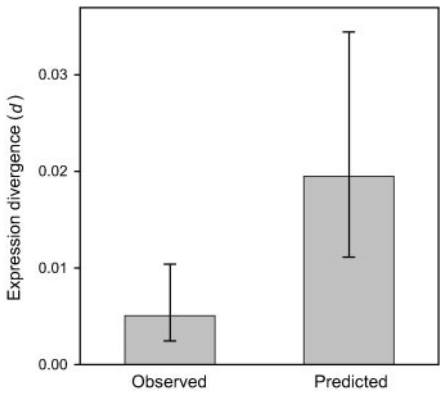

Fig. 2. Predicted and observed divergence (*d*) in gene-expression profiles between two independently evolved clones. The predicted value was calculated as the sum of each clone's divergence from its ancestor. Error bars are 95% confidence intervals.

these divergence values are significantly greater than the baseline ancestral value ($P = 0.0013$ and 0.0171 for Ara+1 and Ara−1, respectively, based on one-tailed *t* tests), indicating the divergence of the profile of each evolved clone from the ancestral state. Finally, Fig. 1*d* compares the expression profiles of the two evolved clones. The two evolved clones are less divergent from one another ($d = 0.0050$) than either is from its ancestor, indicating that the evolved changes in their expression have been largely parallel. Notice especially that many genes that stood out conspicuously as having reduced expression in the comparisons with the ancestors do not stand out in the comparison between the two independently evolved clones, indicating parallel reductions in expression.

To test whether the observed divergence between the evolved clones was significantly less than if their expression changes had been completely independent, we calculated the predicted divergence as the sum of the divergences between each evolved clone and its ancestor. The jackknife method was then used to compute confidence intervals for both the observed and expected divergence levels (24, 25). As shown in Fig. 2, the predicted divergence is significantly greater than the observed divergence as judged by the fact that the 95% confidence interval around each value does not overlap the other value.

Analysis of Significant Changes. To examine candidate genes that may have contributed to adaptive evolution, we sought to identify those particular genes with expression that had changed in parallel in both populations. The large data sets obtained using arrays present the statistical challenge of finding meaningful signals among a background of potential noise. Reliance on arbitrary cutoffs using proportional (fold) changes may suffer from either false positives or low power depending on the level of replication and the extent of variation among replicates (27–29). Using a formal statistical method such as the *t* test to compare expression of each gene between two genotypes or treatments (each based on several replicate measures) avoids arbitrary cutoffs but is likely to yield many false positives. For example, with 4,290 genes in the *E. coli* macroarrays, one expects hundreds of false positives (in the absence of any real effect) at the 0.05 significance level. Conversely, traditional methods such as the sequential Bonferroni correction (30) used to adjust for performing multiple tests comparisons are extremely conservative when there are a great many tests, each involving few replicates; these methods therefore may reject many real differences. To circumvent these problems, we focused on those genes that show significant expression changes in both independently evolved populations, because the number expected to exhibit a false-positive response in two independent cases is much smaller ($0.05^2 \times 4{,}290 \approx 11$). In applying this test we did not specify that the direction of change must be the same in both cases. Thus, any remaining false positives should be approximately balanced between cases in which both populations changed in parallel and cases in which they changed in opposite directions. The degree of deviation from this null expectation can be used to assess the extent of true parallel changes in expression.

Fifty-nine genes fulfilled the stringent criterion of significant expression changes in both evolved clones (Table 1). Remarkably, all 59 underwent parallel changes in the two independently evolved populations (i.e., in the same direction with respect to the ancestral expression level). Of these, 12 genes showed increased expression and 47 had reduced expression relative to the ancestor. The likelihood of such coincident directionality by chance alone is extremely low (Fisher's exact test, $P < 10^{-12}$). Hence, these changes are almost certainly a biologically meaningful reflection of the adaptation that occurred in these populations. We emphasize that because of the coordinated regulation of multiple genes, this parallelism does not imply that all 59 of these genes incorporated separate mutations, nor does it imply that the resulting changes in expression were adaptive in every case. However, the extreme parallelism does strongly indicate that one or more mutations that underlie this suite of changes are responsible for some of the improved fitness during the evolution experiment.

The expression levels of many of these 59 genes are known to be regulated by specific effectors including guanosine tetraphosphate (ppGpp) and cAMP-cAMP receptor protein (CRP). These two effectors control more than half of the parallel-responding genes that have been functionally characterized (Table 1). The cellular concentration of ppGpp is controlled largely by the products of *relA* and *spoT* (31). The concentration of cAMP-CRP depends on the products of *cyaA* and *crp*, which in turn are also influenced by ppGpp (32). Thus, these four genes are strong candidates for harboring beneficial mutations in the evolved clones. Neither 20,000-generation clone has a mutation anywhere in *crp*, *cyaA*, or *relA*. However, the Ara−1 clone has a mutation in *spoT*, in which a change from A to T at the second base of codon 662 causes the replacement of a lysine by an isoleucine.

Effect of the *spoT* Mutation. To isolate the effect of the *spoT* mutation from others that occurred in 20,000 generations, we performed an allelic replacement to move it into the ancestral genetic background. We then competed the ancestral genotype with the *spoT* mutation against the ancestor with the ancestral allele under the same culture conditions that prevailed during the long-term evolution experiment. The *spoT* mutation was beneficial, producing a fitness gain of 9.4% ($t_s = 13.176$, 10 df, $P < 0.0001$). This advantage was eliminated when the *spoT* mutation was replaced again by the ancestral allele, demonstrating that the fitness benefit was caused by the *spoT* mutation and not some artifact of strain construction. Based on growth kinetics in pure culture, the *spoT* mutation confers its advantage both by reducing the duration of the lag phase before exponential growth and increasing the maximum growth rate (data not shown). The 9.4% fitness gain associated with this mutation represents a substantial fraction of the 67% gain measured in this population after 20,000 generations (20).

The benefit of the *spoT* mutation corresponds in magnitude to three step-like gains of ≈8–12% that were seen in the Ara−1 population during the first 2,000 generations of the experiment (19). To examine which of these adaptive steps was associated with this mutation, we performed a PCR/restriction fragment length polymorphism analysis of clones isolated at 500, 1,000, and 1,500 generations to measure the spread of the *spoT* mutant. The mutation was found in 0/100 clones at 500 generations,

Table 1. Genes with significantly changed expression in both independently evolved clones

Gene	Functional group†	Fold change*			Effectors‡
		Ara + 1	Ara − 1	*spoT*	
alr	Amino acid biosynthesis and metabolism	−1.94	−1.97	**−1.24**	ppGpp
dadA	Amino acid biosynthesis and metabolism	−1.63	−2.15	1.20	ppGpp and CRP-cAMP
pheA	Amino acid biosynthesis and metabolism	−1.69	−1.49	**−1.40**	ppGpp
thrC	Amino acid biosynthesis and metabolism	−2.35	−2.40	1.01	ppGpp and CRP-cAMP
hemA	Biosynthesis of cofactors, prosthetic groups, and carriers	−1.77	−1.65	1.05	
menD	Biosynthesis of cofactors, prosthetic groups, and carriers	−1.41	−1.70	1.02	
malP	Carbon compound catabolism	−2.37	−2.03	1.04	
malQ	Carbon compound catabolism	−3.04	−2.58	1.08	CRP-cAMP
rbsK	Carbon compound catabolism	−2.14	−1.92	1.30	CRP-cAMP
flaG	Cell processes (incl. adaptation, protection)	−9.44	−11.74	**−1.29**	CRP-cAMP
flgB	Cell processes (incl. adaptation, protection)	−21.48	−16.36	**−1.56**	CRP-cAMP
flgC	Cell processes (incl. adaptation, protection)	−7.37	−4.74	**−1.28**	CRP-cAMP
flgD	Cell processes (incl. adaptation, protection)	−11.35	−6.78	**−1.55**	CRP-cAMP
flgE	Cell processes (incl. adaptation, protection)	−16.40	−13.40	**−1.94**	CRP-cAMP
flgF	Cell processes (incl. adaptation, protection)	−4.38	−2.82	**−1.46**	CRP-cAMP
flgG	Cell processes (incl. adaptation, protection)	−11.29	−9.61	**−1.77**	CRP-cAMP
flgK	Cell processes (incl. adaptation, protection)	−12.71	−8.36	−1.03	CRP-cAMP
flgL	Cell processes (incl. adaptation, protection)	−6.43	−6.34	−1.15	CRP-cAMP
trg	Cell processes (incl. adaptation, protection)	−2.83	−2.35	1.05	CRP-cAMP
flhA	Cell structure	−1.55	−1.74	−1.03	CRP-cAMP
b1009	Central intermediary metabolism	−1.08	−1.22	1.01	
ccmC	Energy metabolism	1.31	1.16	1.04	FNR
cdsA	Fatty acid and phospholipid metabolism	−1.54	−1.44	1.31	
ybaC	Fatty acid and phospholipid metabolism	−1.56	−2.89	1.40	
yaeT	Hypothetical, unclassified, unknown	−1.50	−1.61	1.15	
yafU	Hypothetical, unclassified, unknown	−1.62	−1.42	−1.39	
ybcU	Hypothetical, unclassified, unknown	−7.00	−3.59	1.08	
b0762	Hypothetical, unclassified, unknown	−1.72	−1.58	**−1.14**	
b1044	Hypothetical, unclassified, unknown	−2.24	−1.72	−1.42	
b1490	Hypothetical, unclassified, unknown	−3.09	−1.54	1.30	
b2445	Hypothetical, unclassified, unknown	1.26	1.28	−1.07	
b2462	Hypothetical, unclassified, unknown	1.29	1.24	1.17	
b2772	Hypothetical, unclassified, unknown	1.29	1.31	1.07	
b2809	Hypothetical, unclassified, unknown	−1.69	−1.21	1.11	
syd	Hypothetical, unclassified, unknown	1.26	1.28	1.01	
yaeJ	Hypothetical, unclassified, unknown	−1.22	−1.39	**−1.13**	
ydfC	Hypothetical, unclassified, unknown	1.20	1.28	−1.03	
ydfE	Hypothetical, unclassified, unknown	1.20	1.25	−1.02	
yjgF	Hypothetical, unclassified, unknown	1.33	1.26	1.04	
yjiD	Hypothetical, unclassified, unknown	1.25	1.19	1.00	
yjjY	Hypothetical, unclassified, unknown	1.24	1.23	1.09	
ykfB	Hypothetical, unclassified, unknown	−1.91	−1.65	−1.09	
yohH	Hypothetical, unclassified, unknown	1.29	1.26	1.15	
nrdA	Nucleotide biosynthesis and metabolism	1.64	1.33	−1.15	Fis
yliJ	Putative enzymes	−3.46	−3.10	1.04	
b1168	Putative enzymes	−1.57	−1.43	1.24	
ynbD	Putative enzymes	−1.40	−1.89	1.32	
yqhC	Putative enzymes	−2.18	−2.14	−1.11	
yhcJ	Putative enzymes	−1.57	−1.76	−1.04	
yidJ	Putative enzymes	−2.32	−1.82	**1.52**	
yjcP	Putative enzymes	−1.64	−1.34	1.11	
hsdR	Transcription, RNA processing, and degradation	−1.91	−1.52	1.16	CRP-cAMP
rpoD	Transcription, RNA processing, and degradation	−1.85	−2.28	1.17	
pepD	Translation, post-translational modification	−2.98	−2.52	−1.01	
malK	Transport and binding proteins	−2.04	−1.90	1.06	
mgtA	Transport and binding proteins	−2.04	−2.14	1.39	
rbsB	Transport and binding proteins	−3.80	−2.53	−1.04	CRP-cAMP
rbsC	Transport and binding proteins	−2.96	−2.08	1.09	CRP-cAMP
rbsD	Transport and binding proteins	−8.35	−7.07	−1.05	CRP-cAMP

*Fold change indicates the proportional change in expression between each 20,000-generation evolved clone and its ancestor and between the ancestral construct that carries the Ara–1 *spoT* mutation and its isogenic counterpart. A value of −2 indicates a halving of expression, whereas a value of +2 indicates a doubling of expression. For the two evolved clones, all changes that have been listed are significant in both at $P < 0.05$. Those changes that are also significant at $P < 0.05$ in the *spoT* mutant are shown in bold type.

†Functional groups are from ref. 36.

‡Effectors are global regulators that control expression of the indicated genes, based on refs. 31 and 41 for this table as a whole and refs. 37 and 42 for *alr* and *pheA*. Note that ppGpp also regulates CRP (32), and thus ppGpp is an indirect effector for all genes shown as being under CRP-cAMP control. FNR, fumurate and nitrate reduction regulatory protein.

41/100 clones at 1,000 generations, and 98/100 clones at 1,500 generations. This pattern does not fit any of the three step-like gains precisely but seems most consistent with the *spoT* mutation contributing to the third adaptive step.

Effect of the *spoT* Mutation on the Ancestral Expression Profile. To confirm that the *spoT* mutation can explain some portion of the expression changes that we saw in the evolved clones, we ran additional arrays to compare the ancestral genotypes bearing the ancestral and derived *spoT* alleles. Of the 59 genes that showed significant changes in both evolved clones, 12 were also altered significantly ($P < 0.05$) by the *spoT* mutation (Table 1). In 11 of these 12, the change was in the same direction as in the evolved clones, which is significantly concordant (binomial test, $P = 0.0032$). Thus, the *spoT* mutation found in the Ara−1 evolved clone produces, by itself, many of the expression changes observed in both populations after 20,000 generations. Notice, however, that the *spoT* mutation does not always account for the full magnitude of the changes; for example, the expression levels of several *flg* genes were reduced 3- to 16-fold in the evolved Ara−1 clone, whereas the *spoT* mutation caused <2-fold reductions. Mutations elsewhere in the genome must cause the other qualitative and quantitative expression changes in the evolved Ara−1 clone. Moreover, there was no *spoT* mutation in the evolved Ara+1 clone, which means that some other mutation is responsible for the changes parallel to those caused by the *spoT* mutation in Ara−1.

In an effort to identify the mutation in Ara+1 causing the parallel suite of changes, we sequenced three other candidate genes: *ptsG* (affecting cAMP-CRP), *gppA* (affecting ppGpp metabolism), and *rpoB* (encoding a target for ppGpp). However, we found no mutations. Although the identity of the mutation causing the parallel suite of expression changes in Ara+1 remains unknown, we tested the hypothesis that this unknown mutation contributed to the same advantageous effect as the *spoT* mutation by moving the *spoT* mutation from Ara−1 to an Ara+1 clone from generation 2,000. The *spoT* mutation conferred no benefit in Ara+1 (mean fitness effect −0.3%, $t_s = 0.1418$, 10 df, $P = 0.8901$). This finding thus confirms our hypothesis that some other beneficial mutation is present in Ara+1 that renders the *spoT* mutation superfluous.

Genetic Parallelism. To examine further the extent of parallel evolutionary changes, we sequenced *spoT* in clones from 10 other populations (besides the two used in the arrays) that also had evolved for 20,000 generations under the same conditions. Seven other evolved populations also had acquired point mutations in *spoT*, causing amino acid replacements in all cases, although no two mutations were identical (Fig. 3). These results indicate that parallelism extended to the gene level in 8 of the 12 populations. An earlier study also uncovered parallel mutations, specifically deletions of varying extent in the *rbs* operon, that affected all 12 populations (8). Four genes in the *rbs* operon were among those showing parallel reductions in expression (Table 1). Thus, expression arrays revealed the effects of previously identified mutations as well as suggested new candidate genes for sequencing and manipulation. In striking contrast to these cases of parallelism in candidate genes, sequencing ≈500 bp in each of 36 randomly chosen gene regions in clones from all 12 populations found only 10 mutations in total, all of them in a few lines that had evolved defects in DNA repair, and no case in which even two populations had mutations in the same gene (33).

Mechanistic Basis for the Effect of the *spoT* Mutation on Fitness. Because ppGpp influences the expression of many genes including those also controlled by cAMP-CRP (32), it is impossible to elucidate the precise physiological basis for the advantage of the *spoT* mutation. However, at least two possibilities can be suggested. First, the array data show that the *spoT* mutation lowers expression of the flagella-encoding *flg* operons (Table 1). The ancestral strain used in the evolution experiment was nonmotile, the selective environment lacked physical structure, and the production of flagella is known to be costly (34). Hence, reducing the expression of these genes could be beneficial (35). Second, a reduction in the concentration of ppGpp, shown to result from mutations in the regulatory region of *spoT* (32, 36), might increase the rate of transcription from tRNA and rRNA promoters (37). This increased transcription raises the maximal growth rate (38), presumably via an increased speed of translation during growth in minimal medium (39). Although tRNA and rRNA genes are not present on the arrays that we used, we can examine the expression of genes with products that are associated with them such as ribosomal proteins and aminoacyl tRNA synthetases. Consistent with this possible advantage, the *spoT* mutation increased the expression of 62 of 94 stable RNA-associated genes when it was moved into the ancestral genome. This proportion is significantly more than half ($P = 0.0027$, one-tailed binomial test).

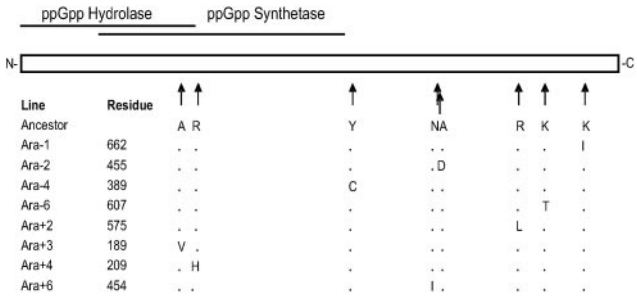

Fig. 3. Nonsynonymous mutations in *spoT* in eight independently evolved *E. coli* populations. Only the variable amino acid residues are shown, with the ancestor listed first and the eight mutant alleles shown below. Four other populations retained the ancestral sequence. The maximum extent of the regions needed for ppGpp hydrolase and synthetase activities are shown along the top; the C terminus is hypothesized to regulate the relative activity of these two functions (36).

Previous work on yeast showed that replicate populations underwent parallel changes in gene expression during 250 generations of evolution (16). We extend this result to another system by demonstrating that expression in independently evolved populations of *E. coli* changed in parallel during 20,000 generations. We then also used these data to gain insight into the genetic basis of adaptation by focusing on the parallel changes, which often indicate important targets of selection. Analysis of these changes led to the discovery of a mutation in the *spoT* gene. This mutation had large phenotypic effects, conferring a very substantial competitive advantage and having widespread pleiotropic effects on the expression of many genes controlled by ppGpp.

The *spoT* mutation was identified as a result of parallel changes in expression in two evolved clones and by itself produced many of these changes when introduced into the ancestral genome. Therefore, the fact that a mutation in *spoT* is present in only one of the two populations used in the array experiments is rather surprising in that light. In our view, the simplest explanation for the absence of a *spoT* mutation in Ara+1 is that one or more mutations in other genes are present in that population that produce a similar suite of phenotypic effects. Consistent with this hypothesis, the *spoT* mutation from Ara−1 does not confer a fitness benefit when moved into Ara+1. The presence of a mutation substituted during the evolution of Ara+1 that renders the effect of the *spoT* mutation redundant represents a case of historical contingency (40). In that respect, it would be interesting to examine whether the order in which

mutations arise might constrain the spectrum of subsequent mutations that can be beneficial.

In conclusion, we used DNA expression arrays to assess the extent of changes in gene expression during 20,000 generations of evolutionary adaptation by *E. coli* to a minimal medium containing glucose as the sole source of energy. Analysis of the expression profiles of two independently evolved lines showed many parallel changes in their gene expression. Concentrating on parallel changes allowed us to reduce greatly the number of false positives that can plague analyses of large data sets and thereby narrow down the number of candidate genes for detailed analysis. These data led to the discovery of a mutation in the *spoT* gene that had large phenotypic effects including a substantial competitive advantage and widespread pleiotropic effects on the expression of many genes controlled by ppGpp. Similar mutations in *spoT* were found in 8 of 12 independently evolved populations. By using parallel expression changes in the discovery of a beneficial mutation, we also demonstrate the utility of viewing gene-expression profiles as phenotypic traits.

We thank F. Moore, S. Remold, and D. Schneider for many valuable discussions; D. Schneider for sharing pDS132; and anonymous reviewers for helpful comments. This work was supported by a grant from the National Science Foundation (to R.E.L.).

1. Nosil, P., Crespi, B. J. & Sandoval, C. P. (2001) *Nature* **417,** 440–443.
2. Harvey, P. H. & Pagel, M. (1991) *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, Oxford).
3. Simpson, G. G. (1953) *The Major Features of Evolution* (Columbia Univ. Press, New York).
4. Bull, J. J., Badgett, M. R., Wichman, H. A., Huelsenbeck, J. P., Hillis, D. M., Gulati, A., Ho, C. & Molineux, I. P. (1997) *Genetics* **147,** 1497–1507.
5. Ferea, T. L., Botstein, D., Brown, P. O. & Rosenzweig, R. F. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9721–9726.
6. Salehi-Ashtiani, K. & Szostak, J. W. (2001) *Nature* **414,** 82–84.
7. Notley-McRobb, L. & Ferenci, T. (1999) *Environ. Microbiol.* **1,** 33–43.
8. Cooper, V. S., Schneider, D., Blot, M. & Lenski, R. E. (2001) *J. Bacteriol.* **183,** 2834–2841.
9. Van Tuinen, M., Butvill, D. B., Kirsch, J. A. W. & Hedges, S. B. (2001) *Proc. R. Soc. London Ser. B* **268,** 1345–1350.
10. Schneider, C. & Moritz, C. (1999) *Philos. Trans. R. Soc. London B* **266,** 191–196.
11. Schneider, C. J., Smith, T. B., Larison, B. & Moritz, C. (2001) *Proc. Natl. Acad. Sci. USA* **96,** 13869–13873.
12. DeRisi, J. L., Iyer, V. R. & Brown, P. O. (1997) *Science* **278,** 680–686.
13. Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O. & Herskowitz, I. (1998) *Science* **282,** 699–705.
14. Arnold, C. N., McElhanon, J., Lee, A., Leonhart, R. & Siegle, D. A. (2001) *J. Bacteriol.* **183,** 2178–2186.
15. Tao, H., Bausch, C., Richmond, C., Blattner, F. R. & Conway, T. (1999) *J. Bacteriol.* **181,** 6425–6440.
16. Rosenzweig, R. F., Sharp, R. R., Treves, D. S. & Adams, J. (1994) *Genetics* **137,** 903–917.
17. Notley-McRobb, L. & Ferenci, T. (1999) *Environ. Microbiol.* **1,** 45–52.
18. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. (1991) *Am. Nat.* **138,** 1315–1341.
19. Lenski, R. E. & Travisano, M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 6808–6814.
20. Cooper, V. S. & Lenski, R. E. (2000) *Nature* **407,** 736–739.
21. Travisano, M. & Lenski, R. E. (1996) *Genetics* **143,** 15–26.
22. Sniegowski, P. D., Gerrish, P. J. & Lenski, R. E. (1997) *Nature* **387,** 703–705.
23. Vasi, F., Travisano, M. & Lenski, R. E. (1994) *Am. Nat.* **144,** 432–456.
24. Sokal, R. R. & Rohlf, F. J. (1981) *Biometry* (Freeman, New York).
25. Lenski, R. E. & Service, P. M. (1982) *Ecology* **63,** 655–662.
26. Donnenberg, M. S. & Kaper, J. B. (1991) *Infect. Immun.* **59,** 4310–4317.
27. Arfin, S. M., Long, A. D., Ito, E. T., Tolleri, L., Riehls, M. M., Paegle, E. S. & Hatfield, G. S. (2000) *J. Biol. Chem.* **275,** 29672–29684.
28. Long, A. D., Mangalam, H. J., Chan, B. Y. P., Tolleri, L., Hatfield, G. W. & Baldi, P. (2001) *J. Biol. Chem.* **276,** 19937–19944.
29. Tusher, V. G., Tibshirani, R. & Chu, G. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 5116–5121.
30. Rice, W. R. (1988) *Evolution (Lawrence, Kans.)* **43,** 223–225.
31. Cashel, M., Gentry, V. J., Hernandez, V. J. & Vinella, D. (1996) in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ed. Neidhardt, F. C. (Am. Soc. Microbiol., Washington, DC), pp. 1458–1496.
32. Johansson, J., Balsalobre, C., Wang, S., Urbonaviciene, J., Jin, D. J., Sonden, B. & Uhlin, B. E. (2000) *Cell* **102,** 475–485.
33. Lenski, R. E., Winkworth, C. L. & Riley, M. A. (2003) *J. Mol. Evol.*, in press.
34. McNabb, R. M. (1996) in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ed. Neidhardt, F. C. (Am. Soc. Microbiol., Washington, DC), pp. 123–145.
35. Edwards, R. J., Sockett, R. E. & Brookfield, J. F. Y. (2002) *Curr. Biol.* **12,** 863–867.
36. Gentry, D. R. & Cashel, M. (1996) *Mol. Microbiol.* **19,** 1373–1384.
37. Barker, M. M., Gaal, T., Josaitis, C. A. & Gourse, R. L. (2001) *J. Mol. Biol.* **305,** 673–688.
38. Sarubbi, E., Rudd, K. E. & Cashel, M. (1988) *Mol. Gen. Genet.* **213,** 214–222.
39. Sorenson, M. A., Jensen, K. F. & Pederson, S. (1994) *J. Mol. Biol.* **236,** 441–454.
40. Travisano, M., Mongold, J. A., Bennett, A. F. & Lenski, R. E. (1995) *Science* **267,** 87–90.
41. Huerta, A. M., Salgado, H., Thieffry, D. & Collado-Vides, J. (1998) *Nucleic Acids Res.* **26,** 55–59.
42. Tedin, K. & Norel, F. (2001) *J. Bacteriol.* **183,** 6184–6196.