Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*

Nadège Philippe,[1] Estelle Crozat,[1] Richard E. Lenski,[2] and Dominique Schneider[1]*

Summary

Evolution has shaped all living organisms on Earth, although many details of this process are shrouded in time. However, it is possible to see, with one's own eyes, evolution as it happens by performing experiments in defined laboratory conditions with microbes that have suitably fast generations. The longest-running microbial evolution experiment was started in 1988, at which time twelve populations were founded by the same strain of *Escherichia coli*. Since then, the populations have been serially propagated and have evolved for tens of thousands of generations in the same environment. The populations show numerous parallel phenotypic changes, and such parallelism is a hallmark of adaptive evolution. Many genetic targets of natural selection have been identified, revealing a high level of genetic parallelism as well. Beneficial mutations affect all levels of gene regulation in the cells including individual genes and operons all the way to global regulatory networks. Of particular interest, two highly interconnected networks—governing DNA superhelicity and the stringent response—have been demonstrated to be deeply involved in the phenotypic and genetic adaptation of these experimental populations.

***BioEssays* 29:846–860, 2007.**

[1]Laboratoire Adaptation et Pathogénie des Micro-organismes, CNRS UMR5163, Université Joseph Fourier, Grenoble, France.
[2]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, USA

Funding agencies: Université Joseph Fourier; Centre National de la Recherche Scientifique; Agence Nationale de la Recherche (ANR); Program "Biologie Systémique (BIOSYS)"; Michigan State University; National Science Foundation; DARPA "FunBio" Program.

*Correspondence to: Dominique Schneider, Laboratoire Adaptation et Pathogénie des Micro-organismes, CNRS UMR5163, Université Joseph Fourier, Grenoble, France. E-mail: dominique.schneider@ujf-grenoble.fr

DOI 10.1002/bies.20629
Published online in Wiley InterScience (www.interscience.wiley.com).

Abbreviations: Ara$^-$, inability to grow on arabinose as a carbon source; Ara$^+$, ability to grow on arabinose as a carbon source; CTD, carboxy terminal domain; IS, insertion sequence; NAD, nicotinamide adenine di-nucleotide; PEP, phosphoenolpyruvate; rDNA, ribosomal DNA; RFLP, restriction fragment length polymorphism; rRNA, ribosomal RNA; tRNA, transfer RNA.

Introduction

Natural selection has shaped all the beautiful and extraordinary diverse phenotypes of living organisms,[1] while random mutations provide the raw material for this process.[2] Although these processes are understood in general terms, it is usually difficult to dissect evolutionary changes in detail because the relevant events happened in the distant past and involved unique circumstances. However, evolution experiments allow one to observe phenotypic and genetic evolution in action.[3] One can even investigate the repeatability of evolutionary outcomes by comparing replicate lineages while they evolve in identical environments.

Micro-organisms are especially attractive for experimental evolution owing to their short generations, which allow experiments to run for hundreds and even thousands of generations, as well as their large populations, which provide a tremendous supply of new mutations available for evolution. Moreover, ancestral and derived organisms sampled during an experiment can be frozen and later revived, so that their phenotypes can be simultaneously measured and directly compared to one another to assess any changes. The use of *Escherichia coli* in evolution experiments offers additional advantages because the tremendous amount of molecular and cellular information available for this model bacterium helps one to analyze and interpret the genetic and mechanistic bases of phenotypic changes that occur during an experiment. As we will show in this review, this wealth of information is particularly valuable for investigating evolved changes in the structure and dynamics of regulatory networks.

In their natural environments, bacteria must survive multiple stresses, and they must cope with repeated bouts of feast and famine. Bacteria sense these environmental changes using complex and highly interconnected regulatory networks, which allow the cells to adjust their gene expression to match the new conditions.[4] These regulatory networks are characterized by a multilayer structure that ranges from local control of individual genes and operons to global control of gene expression across the entire genome.[5] The highest level of regulation must integrate all the environmental cues, and transmit the integrated signal so as to express one set of genes, while repressing others, in order to generate an appropriate response to the current environmental conditions.

The dynamics of the interconnected networks involving these global regulators are therefore critical for bacterial adaptation across both physiological and evolutionary timescales. While the various levels of gene expression have been widely studied from a molecular perspective, the evolutionary dynamics of regulatory networks are largely unexplored. To understand this aspect of evolution, it is necessary to relate changes in regulatory networks to changes in the performance of organisms including their fitness, which provides an integrated measure of the relative survival and reproductive rate of genotypes in a given environment. By linking changes in regulatory networks to changes in fitness, we can investigate how natural selection is able to re-shape and improve these networks, and which regulatory levels are most plastic over evolutionary time. While the focus of our paper is on bacterial evolution, changes in regulatory networks are as, if not even more, important in the evolution of multicellular organisms. For example, the high genetic similarity between such species as humans and chimpanzees—whose orthologous proteins differ, on average, by only two amino acids[6]—has led researchers to emphasize the evolution of regulatory networks in explaining their conspicuous morphological and behavioural differences.[7]

In this review, we present evidence obtained during a long-term evolution experiment with *E. coli* that two networks—those governing the stringent response[8] and DNA supercoiling[9]—constitute key players in these global regulatory dynamics.

Adversity and diversity in nature and the laboratory

Bacteria have evolved for billions of years, leading to extraordinary diversity, for example, in their mechanisms for coping with adverse environment. Their adaptive responses are often manifest as physiological and morphological changes. At one extreme, some bacteria undergo differentiation to yield specialized spores that can withstand starvation and other potentially lethal environments. Some bacteria can form spores at the end of a complex developmental program,[10,11] while others have evolved a multicellular social life that promotes sporulation and survival.[12] At the other extreme, many enteric bacteria undergo subtle morphological and physiological changes when nutrients are depleted, including reduced cell size, modified cell shape, nucleoid compaction, and alterations in cell wall composition and cytoplasmic constituents.[13] Some bacterial pathogens alternate between two phenotypically different states, an infectious but non-replicating form that allows transmission to a new host and a non-infectious but actively replicating form that spreads inside the host.[14,15] The timescales over which these diverse responses to adverse conditions have evolved are largely unknown, but undoubtedly they have been very long.

In this review, we will focus on an evolution experiment performed in the laboratory, one in which *E. coli* populations have experienced transitions between growth and starvation every day since 1988.[16] While this timescale is long for an experiment, we must also emphasize that it is short from an evolutionary perspective, certainly in relation to the diversification of survival strategies described above. However, by virtue of its accessibility and tractability, this experiment provides unique material for relating phenotypic and genomic evolution, including how the bacteria have adapted evolutionarily to the repeated transitions between growth and starvation. After providing a brief overview of the experimental setup, we will describe some of the phenotypic and genetic changes that have evolved, emphasizing the beneficial mutations found in regulatory networks that allow the bacteria to adapt to their fluctuating environment.

The long-term evolution experiment

E. coli B was the ancestral strain used in the long-term evolution experiment.[16] Twelve populations have been propagated by daily serial transfer, using a 1:100 dilution, in the same defined environment for more than 40,000 generations.[17–19] The growth medium is Davis Minimal medium supplemented with glucose at 25 μg/mL (DM25), and that concentration limits the stationary-phase density to $\sim 5 \times 10^7$ cells per mL. Six populations (designated Ara−1 to Ara−6) were founded by the original ancestor, which is unable to use arabinose as a sole carbon source (Ara^-), while the other six (Ara+1 to Ara+6) were founded from an Ara^+, but otherwise isogenic, mutant of the ancestor. The arabinose-utilization phenotype serves as a marker in competition experiments that are performed to measure relative fitness, and it is selectively neutral under these conditions.[17] During each daily cycle, the evolving cells experience the transition from starvation to growth upon transfer into fresh medium, and then back to starvation once the glucose has been exhausted by the growing population. Therefore, selection can act on mutations that affect physiology and performance during lag, exponential, and stationary phases. Under those conditions, the total size of each population fluctuates daily between about 5×10^6 and 5×10^8 cells, given the culture volume of 10 mL.

Genetically heterogeneous mixtures have been sampled from each population and frozen at regular intervals during the experiment, providing a frozen fossil record of sorts. Unlike a typical fossil record, however, the frozen cells can be revived, clones can be isolated, and their phenotypic and genomic attributes can be simultaneously measured and compared with the properties of their ancestor. For example, one can directly compete derived and ancestral bacteria against one another in order to measure their relative fitness, after each type has been acclimated to the same environment. Such competitions can be run under the same conditions as used in the long-term evolution experiment, thus allowing the extent of

improvement to be quantified. Competitions can also be performed under other conditions in order to examine correlated changes in performance capacity.

Phenotypic evolution in the selection environment

Changes in fitness were measured by directly competing evolved populations or individual clones with the ancestor carrying the opposite arabinose marker in the same glucose-limited environment used for the evolution experiment itself.[17] Fitness is calculated as the ratio of the realized (net) population growth rates obtained for two strains while they compete with one another. All the replicate populations achieved substantial fitness gains, averaging about 70% after 20,000 generations.[19] The rate of fitness increase was very fast during the first 2,000 generations, and became progressively slower as the generations elapsed (Fig. 1A).

Several other traits also evolved in parallel in most or all of the twelve populations. The bacteria experienced repeated cycles of growth and starvation, and the evolved populations consistently achieved their higher fitness by growing faster during the exponential phase and by shortening the duration of the lag phase prior to re-commencing growth.[20,21] The most-visual phenotypic change is that average cell volume increased substantially in all twelve populations,[22] with an evolutionary trajectory similar in form to the one for fitness (Fig. 1B).

Phenotypic changes seen in other environments

The bacteria have clearly become much more fit under the glucose-limited, thermally benign conditions where they evolved, but their improvement may not extend to other environments. Indeed, long-term adaptation to the same conditions, day after day, may have generated trade-offs that would be manifest as reduced fitness in other environments, a phenomenon known as ecological specialization. To examine this possibility, the growth rate and yield of the evolved populations and the ancestor were quantified on a wide range of substrates, including 64 that were informative.[19] The resulting data showed numerous significant and parallel declines in the catabolic breadth of the evolved bacteria, indicative of increased resource specialization. In most cases, these losses of physiological capacity were subtle and quantitative, for example a 5% reduction in growth rate or a 10% reduction in yield. In one case, however, a complete loss of the ability to grow was observed for all 12 lines with ribose as the sole carbon source.

Two distinct population-genetic mechanisms could account for the evolutionary trend toward increased ecological specialization. One mechanism is called antagonistic pleiotropy, and it posits that the very same mutations that are beneficial in the selection environment have detrimental effects on performance under other conditions. The other mechanism is called mutation accumulation; it depends on the spread by random drift of neutral mutations in genes that encode pathways that are not being used in the selection environment. Several indirect lines of evidence supported the hypothesis that antagonistic pleiotropy was the main force leading to ecological specialization, including the observation of parallel changes in catabolic function across the replicate lines, most notably the example of ribose noted above.[19] This interpretation was confirmed by the discovery and demonstration of beneficial mutations that confer higher fitness in the glucose environment, but lower fitness measured on other sugars (see also section "Parallel changes in global expression profiles").

One particularly interesting, and dynamically complex, example involves the parallel losses of the ribose catabolic

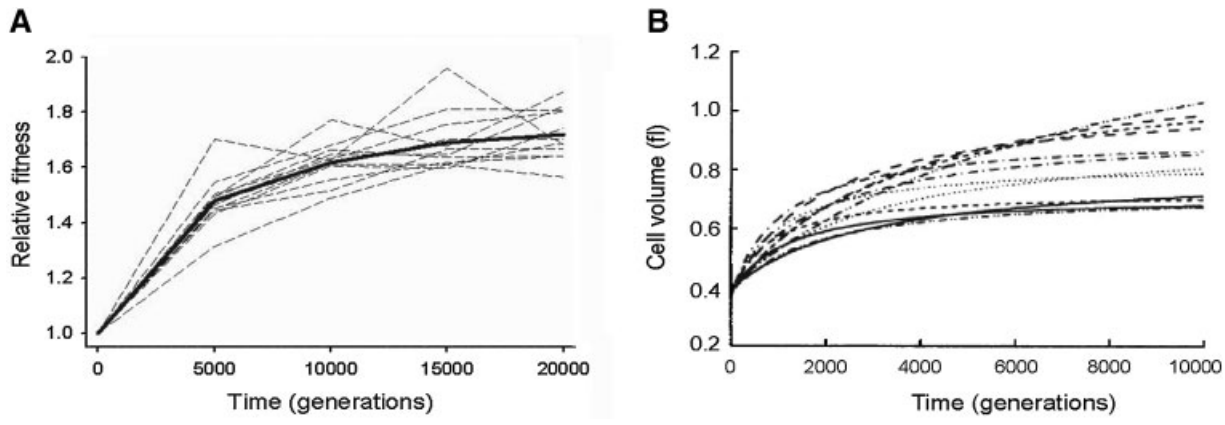


Figure 1. **A**: Fitness trajectories during 20,000 generations of experimental evolution with *E. coli*.[19] Each dotted curve shows the relative fitness values of an evolving population relative to the common ancestor. The solid curve is the mean trajectory of the twelve populations. **B**: Cell size trajectories showing the average cell volume (fL) in each population during 10,000 generations.[18] (Panel B: Copyright 1994 by National Academy of Sciences, USA).

function seen in all twelve lineages.[23] These phenotypic changes resulted from parallel deletions of the ribose-utilization genes, which are clustered in the *rbs* operon. The deletion mutations involved transpositions of IS*150* elements into the *rbs* operon, which were followed by recombination between the resulting new copy and an IS*150* element that was already present upstream of *rbs* in the ancestral strain.[23] These IS element activities contributed to the instability of the *rbs* locus, but selection was also required to explain the rapid spread of the deletion mutants to fixation in all 12 populations. Isogenic strains were constructed that differed only in the presence or absence of the *rbs* operon, and the deletion mutation conferred an advantage of ~2% when these strains competed in the glucose-limited environment.[23] The precise physiological basis for the benefit associated with losing the *rbs* operon is not known, but it might reflect the cost associated with poorly regulated expression of that operon, which is unnecessary for growth on glucose.[24]

Many mutations, few substitutions

A simple calculation allows one to estimate the total number of mutational events that have happened in a typical population as the product of the effective population size (adjusted for serial dilution, 3×10^7), the number of generations (40,000), the mutation rate (5×10^{-10} per bp[25]), and the genome size (4.7×10^6 bp).[16] The total number of mutations during 40,000 generations in each population is thus estimated to be about 3×10^9. Therefore, to a first approximation, all possible mutations at each genomic position were available to the evolving populations. However, most combinations of mutations have not been tested because new mutations generally occurred against the prevailing background, which changed over time.

Based on estimates of the combined effects of natural selection and random drift, probably fewer that 100 mutations have been substituted in a typical population (see Ref. [16] for a full explanation of this estimation). In four populations, however, the expected numbers of mutations and substitutions are higher because the populations evolved mutator phenotypes during the experiment.[19,26] The point-mutation rates in these four populations have increased by ~100-fold. However, that increased mutation rate does not imply a corresponding increase in adaptive substitutions because clones that carry different beneficial mutations compete and interfere with one another in large asexual populations.[27] Three of the mutator populations were shown to be defective in the methyl-directed mismatch repair pathway,[26] and mutations were subsequently found in the *mutL* and *mutS* genes.[28] Mutations that have been found in the fourth mutator population have a sequence signature that suggests that population has a defect in *mutT*, although that has not been directly confirmed.[29] There is no evidence that the mutator mutations themselves confer any direct fitness advantage.[28] Instead, it appears that they have spread by hitchhiking with beneficial mutations.[16,26,27]

Chasing beneficial mutations

Several approaches have been pursued to find beneficial mutations that were substituted in the long-term experimental lines. One approach relies on changes in phenotypic traits that suggest particular candidate genes, as discussed previously for the losses of ribose-catabolic function. However, competitive fitness is an extremely complex trait, and it is not obvious which genes are appropriate candidates for mutations that increase fitness. Thus, a second approach was tried, which involved sequencing a set of 36 randomly chosen genes in the ancestor and in multiple clones sampled at generations 10,000 and 20,000 from all 12 populations.[29] Only 10 mutations in total were discovered this way, and all were found in the mutator populations, even though mutators and non-mutators reached similar fitness levels. Moreover, some of the mutations were synonymous mutations, while others were polymorphic and had not been substituted. These patterns are not what one would expect for beneficial mutations, but instead they reflect non-adaptive mutations that invariably occur, especially at the elevated mutation rates in the mutator populations. These data thus highlight the difficulty of finding beneficial mutations. Nonetheless, the data are useful because they provide a sort of negative control that shows what can be expected at loci that are not targets of selection in the experimental environment.

A third strategy for finding mutations proved to be more successful, and it involved three sequential steps. First, Insertion Sequence (IS) elements[30] were used as genetic markers to follow changes in genotypic diversity over time in two of the evolving populations.[31] IS elements are mobile elements that encode the information required for their own transposition into other genomic locations. They promote mutations through their transposition activity and, when they are present in multiple copies, by serving as substrates for recombination, thereby generating chromosomal rearrangements.[32] IS elements are often used as markers for species typing and for epidemiological source tracking, because they generate many mutations and those mutations are conspicuous from a molecular perspective. A subset of the IS-associated marker mutations observed in the two long-term lines were eventually substituted in their respective populations. The second step, therefore, was to characterize the molecular events responsible for these IS-associated substitutions and identify the affected genes.[33] The third step in this strategy was to sequence some of the same loci in which new IS insertions were substituted in one population in clones sampled from the other eleven populations. This step was performed for four genes, and led to the discovery of 36 additional mutations that showed several striking patterns indicative of beneficial mutations.[34] In particular, they were mostly non-synonymous point mutations, and they were

spread uniformly across the non-mutator and mutator lineages. Moreover, the number of mutations found in these genes was completely different from that seen for the randomly chosen genes. Such parallelism across multiple lineages has long been recognized as evidence for adaptive evolution.[35,36]

Two additional strategies have also proven to be productive for finding mutations—especially ones that affect gene-regulatory networks—in these evolving lines. One approach has been to compare global gene-expression profiles of evolved and ancestral clones using transcriptomic and proteomic methods.[37,38] The other has been to screen for changes in phenotypes, such as DNA topology, that are known to have subtle yet widespread affects on gene regulation and expression.[39] The important regulatory mutations found by all these approaches will be described in the sections that follow.

Genetic diversity and parallelism

IS fingerprints of evolving populations

The genetic diversity was tracked over the first 10,000 generations in two of the populations by performing restriction fragment length polymorphism (RFLP) analyses using the seven IS elements known in *E. coli* B as molecular probes.[31] Many clones were isolated at several time points from the two focal populations, called Ara−1 and Ara+1, and their IS genomic fingerprints were compared with that of the ancestor. This approach revealed substantial genotypic diversity, such that almost all individuals tested in each population at generation 10,000 were different from one another (Fig. 2). This pattern contrasts with the failure to find any point mutations in 36 randomly chosen genes in any of the eight non-mutator populations after 20,000 generations,[29] which shows the utility of IS markers for distinguishing closely related strains. Of course, not all IS-associated mutations are beneficial. Phylogenetic trees were constructed based on the IS fingerprints, which allowed a subset of mutations to be identified that had been substituted in their respective populations (Fig. 2). These "pivotal" mutations represent good candidates for conferring higher fitness, although their substitution could also have occurred by hitchhiking along with beneficial mutations in other genes.

Characterization of IS-mediated mutations

Several of these pivotal mutations were studied in detail in order to characterize the nature and location of the mutations. In addition to the *rbs* deletions described earlier, nine other IS-mediated mutations were characterized.[33] Two types of mutations were discovered: new insertions of IS elements following transposition events, and chromosomal rearrangements after recombination between homologous IS copies. Both focal populations had acquired a large chromosomal inversion; one involved two IS*1* elements, and the other involved two IS*150* elements. One population also had an additional IS-associated deletion. The other six mutations were new insertions of IS elements. Table 1 summarizes information on four of these insertions, each of which was substituted in its respective population during the first 2,000 generations, when the rate of fitness improvement was fastest. The genes that harbored these four early insertions were then sequenced in clones from all 12 populations, in order to assess whether parallel substitutions had occurred in the other lineages, as would be expected if these genes were indeed targets of selection.[34]

Population Ara−1 had a transposition of IS*150* into the *pykF* gene, which encodes pyruvate kinase I. This enzyme is one of two glycolytic pyruvate kinases in *E. coli* that catalyze the transformation of phosphoenolpyruvate (PEP) into pyruvate,[40] and the insertion potentially leads to its inactivation. Population Ara+1 substituted three IS*150* transposition mutations. One IS*150* element jumped 11-bp upstream of the *pbpA-rodA* operon, encoding penicillin-binding protein 2 and the morphogenic protein RodA, both of which are involved in cell wall biosynthesis and cell shape determination.[41] The two other insertions were in the coding sequences of *nadR* and *hokB-sokB*. The *nadR* gene encodes a regulatory protein involved in NAD metabolism; that protein both represses the NAD biosynthetic genes and influences the transport of NAD precursors.[42,43] The *hokB-sokB* locus is homologous to addiction modules that are involved in maintaining plasmids inside bacterial cells.[44]

Genetic parallelism across evolving populations

As predicted under the hypothesis that these new insertions were beneficial mutations, these same four loci harbored substitutions in many or even all of the other independently evolved lineages after 20,000 generations.[34] Mutations were substituted in both *pykF* and *nadR* in all 12 populations, while several populations had mutations in *pbpA-rodA* and *hokB-sokB*. This pattern is in striking contrast to the data obtained by sequencing 36 random genes, where no mutations were found in any of the eight non-mutator populations and no gene was affected in more than one population.[29] Several statistical patterns strongly support the conclusion that selection drove the parallel changes in these four genes, and that the mutations are therefore beneficial in the environment of the long-term evolution experiment.

The functional bases of the benefits conferred by the mutations in these four genes remains the subject of on-going research, but plausible hypotheses can be advanced.[33,34] The *pykF* mutations probably slow down the conversion of PEP to pyruvate. While this effect might seem disadvantageous at first glance, it should lead to an increased concentration of PEP, which provides the energy source that the phosphotransferase system uses to bring the limiting glucose into the cell. The *nadR* mutations probably decrease the NadR repressor activity and thereby increase the intracellular level of

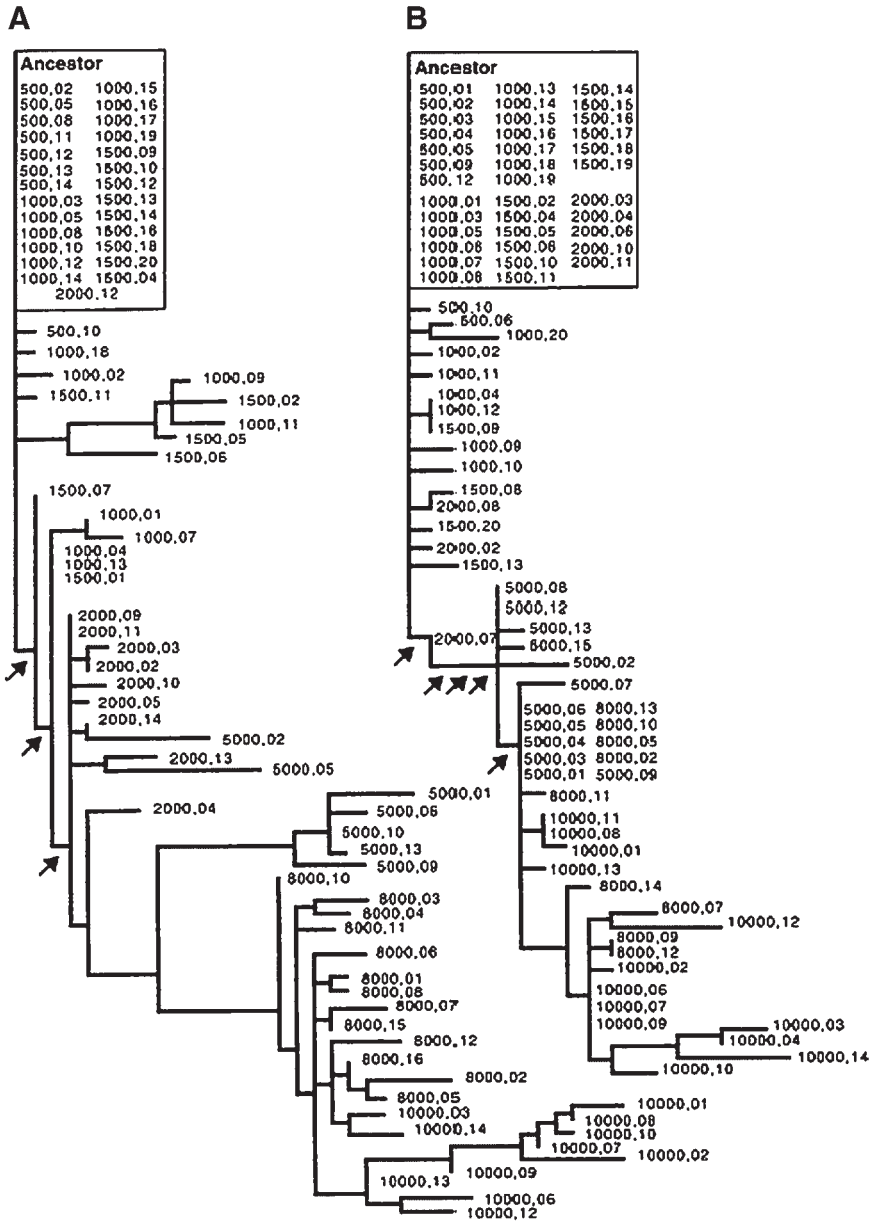


Figure 2. Phylogenetic trees constructed from RFLP-IS analyses of the ancestor and evolved clones sampled from two focal populations during 10,000 generations.[31] Each individual clone is labelled by the generation at which it was sampled followed by an arbitrary number. Clones shown in the boxes at top had the same genetic fingerprints as the ancestor. Mutations that were substituted in the evolving population are marked with arrows. **A:** Population Ara+1. **B:** Population Ara−1. (Copyright 1999 by National Academy of Sciences, USA).

NAD, which may allow the evolved cells to grow faster during exponential phase or might facilitate the daily transition between starvation and growth. Inactivation of the *hokB-sokB* locus may be beneficial to the cells because it removes a toxin–antitoxin system that is not only unnecessary in the absence of plasmids, but also costly in terms of wasted production and perhaps some residual toxicity. Lastly, the mutations in the *pbpA-rodA* operon may change the levels or activities of these two proteins, which are involved in coupling cell wall biosynthesis to cell division. Changes in this coupling process may well contribute to the larger cell volume that evolved in all the populations, as described earlier. We also point out that three of these evolving genes affect key processes and metabolites—PEP and the phosphotransferase

Table 1. IS insertion mutations substituted during the first 2,000 generations of experimental evolution in one of two focal populations

Mutational event	Gene function	Focal population*	Other populations**
Insertion of IS*150* into *pykF*	Pyruvate kinase	Ara−1	11
Insertion of IS*150* into *nadR*	NAD metabolism	Ara+1	11
Insertion of IS*150* into *hokB-sokB*	Plasmid addiction module	Ara+1	5
Insertion of IS*150* upstream of *pbpA-rodA*	Cell wall biosynthesis	Ara+1	5

*Indicates population in which the initial mutation was found by RFLP-IS fingerprinting, and then characterized at the genetic level to identify the affected locus[33].
**The number of other populations (excluding the focal population), out of 11 maximum, in which sequencing the corresponding gene revealed additional mutations[34].

system (albeit indirectly), NAD, and cell division—such that these mutations may impinge on global gene-regulatory networks in many ways.

Genetic diversity within evolving populations

Substitutions were often parallel across the replicate populations, demonstrating the power of selection to find repeatedly the same genetic targets. But this parallelism does not preclude diversity within a given population. Indeed, IS fingerprints revealed extensive genotypic diversity in each focal populations.[31] These data also showed an interesting discordance over time between rates of phenotypic and genomic change. While the rates of change in fitness and cell size decreased over time, this trend was not observed for genomic change. Instead, genetic diversity within each population showed alternating bouts of increase and decline, which provide a signature of so-called "periodic selection" events. These events occur in asexual populations whenever a beneficial mutation sweeps to fixation, transiently purging diversity as the clone bearing that beneficial mutation takes over the entire population. As that clone spreads, it too begins to accumulate genetic variability, which increases until the next beneficial mutation sweeps to fixation. Over time, as the experimental populations become better adapted to their environment, the opportunity for further improvement is reduced, and mutations with smaller fitness advantages become more important to the overall dynamic. In fact, several beneficial mutations may accumulate in a single clone before it can achieve a sufficient advantage to exclude other clones that also carry one or more beneficial mutations. In that case, the rate of genomic evolution may decline slowly compared to the rate of fitness improvement because more mutations are required to produce smaller gains.

Evolution of global regulatory networks

Parallel changes in DNA supercoiling

The transitions between growth and starvation are known to affect the DNA supercoiling level in bacteria, presumably allowing cells to modulate gene expression in ways that facilitate their growth and survival.[5,45] All the populations in the long-term experiment experienced regular and predictable daily transitions between growth and starvation. Hence, it is interesting to ask whether they evolved any changes in their DNA supercoiling state. To measure this trait, a reporter plasmid was introduced into three clones isolated from each population at each of three time points: 2,000, 10,000 and 20,000 generations.[39] The data showed that 10 of the 12 populations evolved greater DNA supercoiling levels, with at least some of the modification already seen by 2,000 generations in most cases. These rapid and parallel changes in DNA supercoiling strongly suggest that the responsible mutations confer higher fitness on the cells.

DNA supercoiling is a tightly controlled process, and two types of enzymes maintain the required superhelicity of the chromosome: topoisomerases[46] and histone-like proteins.[47] To date, the corresponding genes have been sequenced in the focal population Ara−1, which showed two successive increases in DNA supercoiling, one before 2,000 generations and the other after 10,000 generations. Mutations were found in both the *topA* and *fis* genes.[39] The extent of genetic parallelism affecting altered supercoiling in the other lines is currently under investigation. The *topA* gene encodes topoisomerase I, which helps relax DNA by removing negative supercoils[48] and couples transcription with translation by removing hypernegative supercoils behind transcribing RNA polymerase molecules.[49] The *fis* gene encodes a histone-like protein, a global regulator of gene expression including ribosomal and tRNA genes, that helps couple the physiological state of cells to their nutritional conditions.[50] The Fis protein adjusts DNA supercoiling by binding to DNA and regulating the topoisomerase-encoding genes (Fig. 3). The effects of Fis on these genes are reported to be complex. Fis can either upregulate or downregulate *topA* depending on its cellular concentration.[51,52] Fis also represses the *gyrAB* genes that encode DNA gyrase, the only enzyme able to introduce negative supercoils into DNA molecules.[53] The temporal dynamics of both evolved mutations were analyzed using samples frozen during the long-term experiment. The new

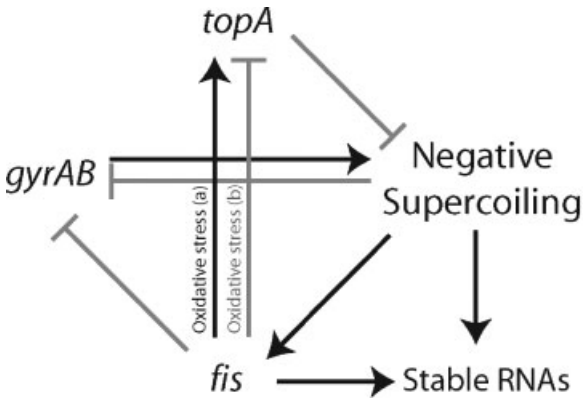


Figure 3. Regulatory interactions involved in adjusting DNA superhelicity. Only those interactions directly relevant for the long-term evolution experiment are shown here. The level of DNA superhelicity is maintained through the combined action of topoisomerases (the *gyrAB* genes encode both subunits of DNA gyrase and *topA* encodes topoisomerase I) and histone-like proteins, including Fis. The Fis protein is a transcriptional regulator of *gyrAB* and *topA*, although its regulation of *topA* is complex. Fis apparently either activates (a) or represses (b) *topA* when Fis is present at low or high concentration, respectively (see text for details). Both Fis and highly negative superhelicity promote transcription of stable RNAs. Pointed arrows indicate activation, while blunt arrows show repression.

topA allele arose before 2,000 generations, whereas the new *fis* allele was substituted after 10,000 generations, and these dynamics are consistent with the two distinct steps in DNA supercoiling seen in this focal population.

A set of isogenic strains was constructed by moving the evolved *topA* and *fis* alleles, alone and in combination, into the ancestral genetic background.[39] Competition assays revealed that both evolved mutations were beneficial. Moreover, they contributed additively to the increased DNA supercoiling, reproducing precisely the net change that was observed in the population. The evolved *topA* allele therefore led to decreased activity of topoisomerase I, and the evolved *fis* mutation was shown to decrease the level of this global regulator as well.[39] The Fis protein activates transcription of *topA* under conditions of oxidative stress[51] and at low Fis concentrations.[52] However, the evolved *fis* allele might reduce the Fis level below that needed to activate *topA*, thereby causing a further reduction in TopA and, consequently, a further increase in DNA supercoiling. These findings demonstrate that DNA supercoiling was an important target of selection during the long-term evolution experiment.

Parallel changes in global expression profiles

Global patterns of gene expression were investigated, at both mRNA and protein levels, by comparing the ancestor with clones sampled at 20,000 generations from focal populations Ara−1 and Ara+1.[37,38] Owing to the two different methodologies and their requirements, slightly different growth conditions were employed. Transcriptional profiles used DNA macroarrays, with total RNAs extracted from cells during exponential growth phase in the same medium as for the long-term evolution experiment.[37] Proteomic profiles used two-dimensional protein electrophoresis, with proteins extracted from cells harvested after stationary phase in the same medium except with ten-fold higher glucose concentration in order to obtain more protein.[38] The macroarrays allowed quantification of expression for essentially all genes in *E. coli*, while the proteomic profile separated about 300 to 400 proteins. Despite these methodological differences, however, the two approaches yielded similar results, with both of them implicating changes in the same global regulatory network.

The transcription profiles revealed significant changes in both independently evolved lines in the expression of 59 genes,[37] while 38 proteins showed significant changes in abundance in both evolved lines.[38] Remarkably, all of the transcriptional and protein changes that were significantly altered in both focal populations were parallel, that is they had evolved in the same direction relative to the ancestral expression. Such parallel evolution is very unlikely to happen by chance; rather, it supports the adaptive importance of these changes. However, it is also unlikely that each gene whose expression changed acquired its own mutation. Instead, these data implicated mutations affecting global regulatory networks that coordinate the expression of large ensembles of genes, along with perhaps some mutations affecting the expression of smaller regulons, individual genes, or both.

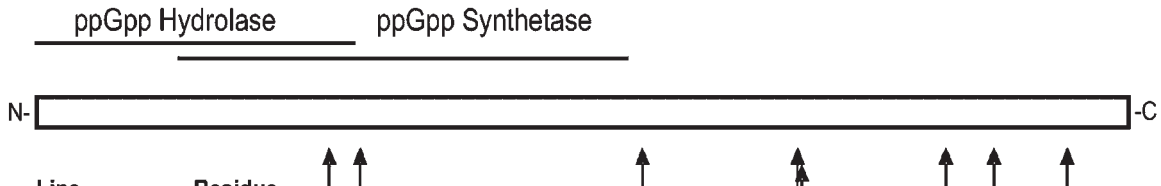
The precise genes where parallel changes in expression were found differed between the transcriptional and proteomic analyses, which is not too surprising because they were performed under different conditions, with the transcriptional analyses performed on exponentially growing cells while proteomic analyses were done using cells in stationary phase.[38] More interestingly, however, a close examination revealed that about half of all the affected genes identified by both approaches belong to the same regulatory network, namely the stringent response that is involved in physiological adaptation to nutritional stress.[8] The effectors of the stringent response are two nucleotide molecules, guanosine tetra- and penta-phosphate (ppGpp and pppGpp, respectively), which are collectively called (p)ppGpp. In the next section, we will give a complete overview of the stringent response regulatory network, together with the changes that occurred during the evolution experiment. For now, it suffices to say that two main genes are known to be involved in eliciting the stringent response and in the metabolism of the effector molecules:[8] *relA* and *spoT*.

Sequencing both of these genes revealed a mutation in the *spoT* gene in one of the two focal populations. When this

evolved *spoT* allele was moved into the ancestral genetic background, it conferred a significant competitive advantage relative to its isogenic counterpart carrying the ancestral *spoT* allele.[37] Moreover, when the global transcriptional profile was analyzed for the constructed ancestral strain carrying the evolved *spoT* allele, it showed many of the changes in gene expression seen in the evolved populations. Interestingly, however, the other focal population, which showed parallel changes in expression, had no mutation in *spoT* nor in any of several other candidate genes known to be involved in the stringent response.[37] Evidently, mutations in some other gene can produce similar effects. Besides the two focal populations, mutations affecting *spoT* were also found in seven of the other ten evolved lines (Fig. 4).[37] It is clear, therefore, that evolutionary adaptation during the long-term experiment has occurred, in part, through mutations affecting the stringent-response regulatory network, with different subsets of genes belonging to this response network affected by the different physiological conditions used in the transcriptomic and proteomic studies.

Besides the global expression changes associated with the stringent-response network, both the proteomic and transcriptomic studies also revealed two smaller regulons whose expression had changed in parallel: the ribose operon and the maltose regulon.[38] As described earlier, ribose catabolic function was lost in all 12 populations, these losses were shown to reflect deletions of large parts of the *rbs* operon, and the deletions were advantageous in the glucose-limited environment.[19,23] Similar work was therefore done to investigate the molecular bases and performance consequences of the altered expression of the maltose regulon. Mutations in *malT*, the transcriptional activator of the maltose operons,[56] were found in both of the focal populations.[38] Isogenic strains were constructed by introducing each evolved *malT* mutation into the ancestral genome. These *malT* mutations reduced growth on maltose but enhanced competitive fitness in the glucose-limited medium used during the evolution experiment.[38] These results, along with similar findings obtained for *rbs* deletions, provide direct evidence supporting the antagonistic-pleiotropy hypothesis, whereby ecological specialization evolves by mutations that enhance fitness on one resource while reducing performance on others. The *malT* gene was also sequenced in the other ten lineages, and mutations were found in six of them, so that eight of the twelve populations had substituted *malT* mutations by 20,000 generations.

Summarizing this section, comparisons of global expression profiles in independently evolved lineages revealed parallel changes in expression that led to the discovery of parallel genetic modifications affecting two different levels of gene regulation: the lower level of operons, and the higher level of global regulatory networks. Mutations affecting the lower level generated straightforward phenotypes, including the declines of ribose and maltose catabolic functions, and the resulting fitness gain in the glucose environment where these functions were evidently costly to maintain. Owing to the potential for widespread pleiotropy affecting many different functions, it is more difficult to predict the phenotypic consequences of mutations affecting the global regulatory



Line	Residue								
Ancestor		A	R	Y	NA		R	K	K
Ara-1	662	.	.	.	.	.	.	.	I
Ara-2	455	.	.	.	.	D	.	.	.
Ara-4	389	.	.	C	.	.	.	.	.
Ara-6	607	.	.	.	.	.	.	T	.
Ara+2	575	.	.	.	.	.	L	.	.
Ara+3	189	V	.	.	.	.	.	.	.
Ara+4	209	.	H	.	.	.	.	.	.
Ara+6	454	.	.	.	I	.	.	.	.

Figure 4. Map showing positions of *spoT* mutations found in eight independently evolved populations.[37] The amino-acid changes are shown together with their locations in the SpoT protein. The SpoT protein has three functional domains: the ppGpp hydrolase domain and the ppGpp synthetase domain,[54] and the carboxy terminal domain (CTD) that may be involved in regulating the other two domains.[55]

level. In the next section, we explore some potential consequences of the *spoT* mutations.

Evolutionary modifications of the stringent-response regulatory network

Overview of the stringent response

The stringent response is a global regulatory network that allows bacteria to cope with their changing environments, in particular with changes in nutritional state.[8] The stringent response is induced by an imbalance between tRNA aminoacylation and protein synthesis, which inhibits the accumulation of stable RNAs upon starvation. Two main lines of evidence first showed that (p)ppGpp is the effector of the stringent response. First, its concentration increases in cells during the transition from growth to starvation.[57] Second, (p)ppGpp is absent in mutants that are defective in the stringent response.[58] More generally, there is an inverse correlation between the concentration of (p)ppGpp and growth rate.[59] In response to starvation for various nutrients, the intracellular level of (p)ppGpp increases via the action of two enzymes, RelA and SpoT, encoded by the *relA* and *spoT* genes, respectively (Fig. 5). The (p)ppGpp effector, through its direct interactions with RNA polymerase and with the help of proteins including DksA,[60] then reprograms the entire transcriptional machinery of the cell. These transcriptional changes inhibit the synthesis of stable RNAs along with various other processes involved in cell growth (Fig. 5). At the same time, (p)ppGpp also stimulates the transcription of certain genes that promote survival during starvation,

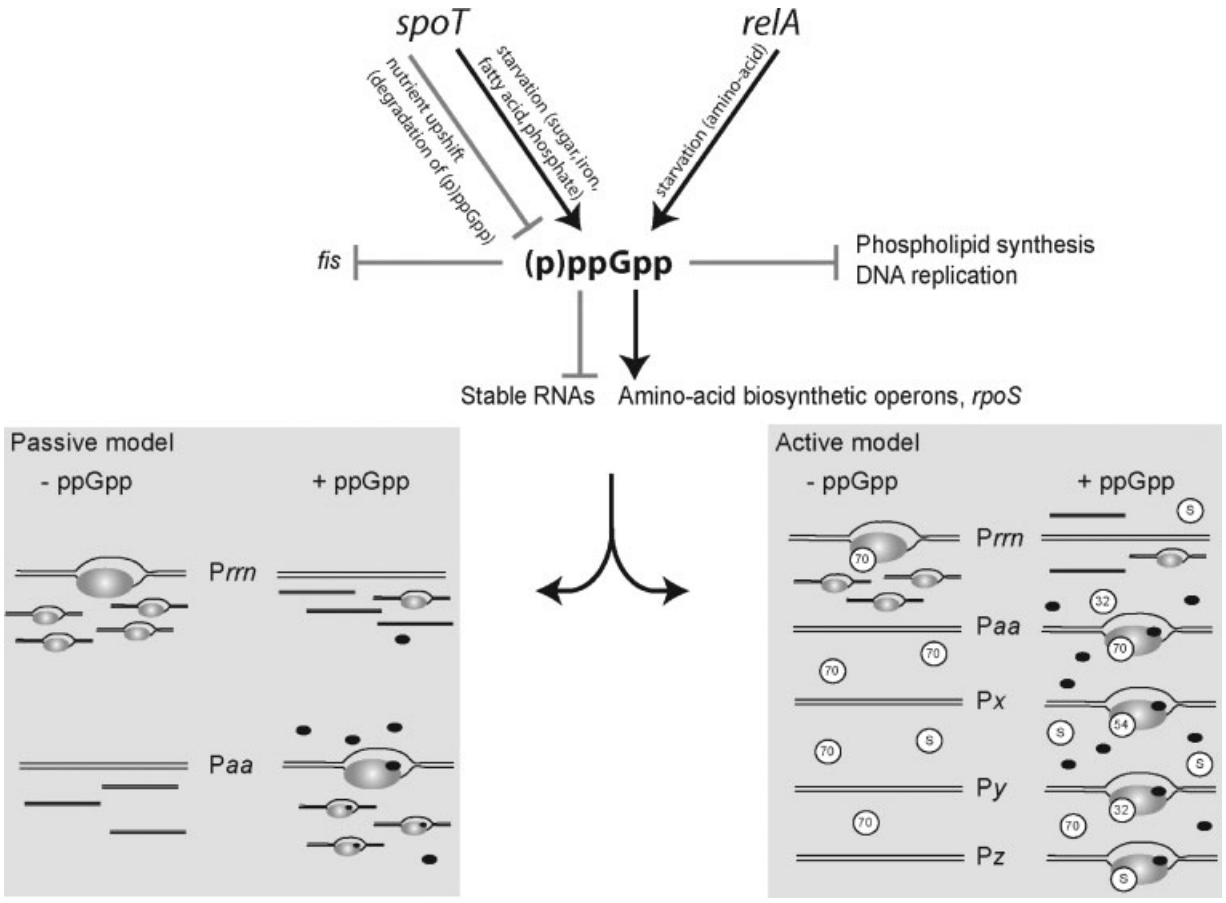


Figure 5. Metabolism of (p)ppGpp and its impact on global transcription. During starvation, two genes are involved in the synthesis of (p)ppGpp, *relA* and *spoT*. When nutrients are present, *spoT* is also involved in degradation of (p)ppGpp. The (p)ppGpp is the effector of the stringent response, which involves substantial reprogramming of the transcriptional machinery in the bacterial cell (see text for details). High levels of (p)ppGpp lead mainly to inhibition (blunt arrows) of transcription of stable RNAs and activation (pointed arrows) of transcription of those genes involved in recovery from nutritional stress. Both passive and active models of gene regulation, as shown in the lower panels, have been proposed to explain these effects (see text for details). Small black ovals represent (p)ppGpp molecules, while the larger grey ovals indicate core RNA polymerase molecules. Circles labelled with 70, S, 32, and 54 indicate the corresponding sigma factors.

including the amino-acid biosynthetic operons and the gene encoding the alternative sigma factor RpoS (Fig. 5).[61] When nutrients become available again, (p)ppGpp is degraded by the SpoT protein,[54] thus allowing transcription of growth-promoting genes to recommence.

The metabolism of (p)ppGpp is largely controlled by two proteins, RelA and SpoT. The RelA protein is a ribosome-associated (p)ppGpp synthetase that responds to amino-acid starvation, with the accumulation of uncharged tRNAs providing the induction signal.[62] SpoT is a bifunctional enzyme with (p)ppGpp synthetase and hydrolase activities localized in the N-terminal half of the protein,[54] followed by a C-terminal domain (CTD) of unknown function (Fig. 4). The CTD has been proposed to be involved with the regulation of both enzymatic activities.[55] While the RelA function is specific to amino-acid starvation, SpoT regulates (p)ppGpp in response to diverse starvation conditions, although the proximate signals leading to (p)ppGpp synthesis or degradation by SpoT are mostly unknown.

Evolution of the stringent response

Recall that eight of the twelve independently evolved lineages substituted mutations in the *spoT* gene (Fig. 4). Six of these mutations affect the CTD of the SpoT protein, and the *spoT* allele from population Ara−1 has been demonstrated to be beneficial.[37] We are currently investigating that mutation in greater detail, along with those substituted in populations Ara+2 and Ara−4 (Fig. 4). We have moved each of these evolved *spoT* alleles into the ancestral chromosome in order to examine their phenotypic effects, and we now briefly summarize some of our findings. Competition assays between these strains and their ancestor, performed in the same environment as used for the evolution experiment, confirm the fitness advantage of the *spoT* allele from population Ara−1, and further show that the two other evolved alleles are also beneficial, with fitness advantages of 10–12% in all cases (our unpublished data).

The simplest explanation for the beneficial effects of these mutations would be that they alter the intracellular concentration of (p)ppGpp by affecting the rates of synthesis, degradation, or both. To test this hypothesis, we performed quantitative assays of (p)ppGpp levels using the same four isogenic strains (the ancestor and three evolved *spoT* alleles in the ancestral background), by nucleotide labeling and one-dimensional thin-layer chromatography.[63] However, there were no measurable differences in accumulation rate, degradation rate, or overall level of (p)ppGpp between the isogenic strains bearing the ancestral and evolved *spoT* alleles (our unpublished data).

Evidently, the fitness benefits conferred by the evolved *spoT* mutations are independent of the intracellular (p)ppGpp levels. The molecular basis for the benefit conferred by the evolved *spoT* mutations therefore remains unknown, although one can imagine various hypotheses. The *spoT* mutation found in population Ara−1 affects the expression of many genes,[37] and some of these changes may increase fitness. For example, this mutation reduces transcription of genes involved in flagella synthesis,[37] and inactivation of flagella synthetic operons in *E. coli* has been shown to increase fitness in well-mixed liquid cultures.[64] Alternatively, recent studies have demonstrated physical interactions between SpoT and various other proteins,[65] including CgtA involved in ribosome maturation[66] and the acyl-carrier protein involved in fatty-acid biosynthesis.[67] The evolved *spoT* alleles might affect these protein interactions, thereby leading to various other changes in the regulatory network of the cell.

In any case, beneficial mutations have been selected in two major regulatory networks, the stringent response and DNA supercoiling, during the long-term evolution experiment. Both networks are involved in regulating cellular responses to growth transitions. The evolved lines were subjected for many years to daily bi-directional transitions between growth and starvation, always in the same glucose-limited medium. There was evidently considerable scope for fine-tuning these coordinated molecular responses to the specific transitions of the long-term experiment. Indeed, the evolved lines have not only increased their maximum growth rates but they have also shortened the lag phase that they experience prior to recommencing growth after being transferred into fresh medium.[20,21]

Evolution of hubs and spokes

Among the regulatory genes targeted by selection during the long-term experiment, some are involved with low-level regulation of individual genes or operons, whereas others encode global regulators that integrate and affect complex networks of interacting genes and their products. This dichotomy reflects the overall architecture of regulatory interactions in many organisms; most genes are "spokes" that have only a few connections to other genes, whereas a few are "hubs" that interact with many other genes.[68] As exemplars of spoke-like genes that have been evolving in the long-term experiment, we would point first to those involved in specific catabolic functions. In particular, deletions of the *rbs* operon and mutations in the *malT* activator gene eliminate or reduce growth on ribose and maltose, respectively. These mutations presumably confer their small fitness advantages in glucose by minimizing wasteful gene expression; they presumably have few, if any, other effects. Other evolving genes that affect cell wall biosynthesis (*pbpA*) and other core metabolic functions (*pykF*, *nadR*) might also be regarded as spokes, although this characterization is less certain in these cases because such core functions may subtly interact with many other genes. For example, *pykF* encodes a kinase that converts PEP and ADP into pyruvate and ATP. But PEP

also drives the transport of glucose into the cell via the phosphotransferase system, and the availability of PEP for this key process may be increased by mutations that reduce the activity of the PykF protein.[33,34] In reality, the distinction between spoke and hub genes is probably graded, without a precise boundary between the classes.

Although certain genes are difficult to place in the spoke and hub framework, some of the genes that changed repeatedly during the long-term experiment must be regarded as hubs because they are central players in two global regulatory networks: DNA supercoiling (*topA, fis*) and the stringent response (*spoT*). Both networks sense environmental fluctuations and induce global changes in gene expression, thereby allowing the cells to acclimate to changing conditions.[5,8,9,60] As noted above, the long-term experiment provides an exquisite training ground for bacteria to adapt to transitions between growth and starvation. Of course, bacteria in nature also face transitions between feast and famine,[69] but this experiment is probably unusual in the temporal predictability and suddenness of the transitions. The rates of synthesis of rRNA and tRNA are proportional to cell growth rate, whereas they decline sharply during the transition into stationary phase (Fig. 6).[70] The genes *spoT*, *topA* and *fis* are all intimately involved with controlling these transitions. In particular, SpoT controls the concentration of (p)ppGpp, which represses transcription from the rDNA promoters as cells enter stationary phase (Fig. 6).[58,71] To exit from stationary phase and recommence growth, DNA superhelicity must be highly negative,[72] and the histone-like Fis protein must be bound to these same rDNA promoters for activation (Fig. 6).[73] It is very striking, therefore, that all three of these genes evolved during the long-term experiment.

A molecular scenario for selection on these regulatory hubs

Increased concentrations of (p)ppGpp have been shown to destabilize transcription initiation complexes, especially the stringent promoters that are negatively regulated by (p)ppGpp, and which are already intrinsically unstable.[74] A crucial determinant of this negative regulation is a specific sequence, called the discriminator, located between the −10 promoter box and the transcription initiation site, and which has high GC content in stringent promoters.[75] These stringent promoters also require highly negative DNA superhelicity for maximal transcription because this GC-rich discriminator sequence provides an energy barrier that is overcome by supercoiling.[76] During growth, superhelicity is highly negative and (p)ppGpp concentration is low, both conditions favoring transcription of rRNA (Fig. 6). During the transition into stationary phase, (p)ppGpp levels rise sharply[57] and the DNA becomes more relaxed,[45] thereby inhibiting the transcription of rRNA (Fig. 6).

During starvation, (p)ppGpp also has a positive effect on transcription of certain other genes, including the amino-acid biosynthetic operons and those genes under the control of alternative sigma factors (Fig. 5).[60] These stimulatory effects have been proposed to result from three mechanisms, one passive[77] and two active.[78,79] During growth, most RNA polymerase molecules are actively engaged in transcribing the stable rDNA operons, but these complexes are destabilized by (p)ppGpp as the cells enter stationary phase (Fig. 5). The passive model for (p)ppGpp-associated activation is thus mediated by the resulting increased availability of RNA polymerase molecules to transcribe promoters for which they have much lower affinity, such as the promoters for the amino-acid biosynthetic operons.[77] From the active standpoint, first,

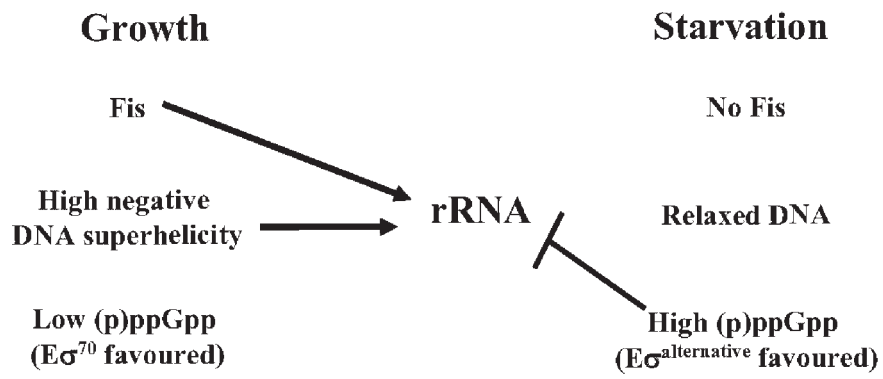


Figure 6. Comparison of the main regulatory interactions differentially affecting ribosomal operons during growth versus starvation. During growth, the combination of a high level of Fis, highly negative DNA superhelicity, and a low level of (p)ppGpp promotes increased transcription of ribosomal operons, whereas starvation leads to the opposite combination of factors (see text). Pointed and blunt arrows represent activation and repression, respectively.

(p)ppGpp has been shown to directly stimulate the transcription initiation step of amino acid promoters, together with the DksA co-regulator protein.[78] The specificity towards amino acid promoters, compared to the stringently controlled promoters, reflects the different kinetic parameters of transcription initiation between these two sets of promoters. Second, (p)ppGpp also mediates the molecular competition between sigma factors for the RNA polymerase core enzyme, which it does by increasing the binding ability of alternative sigma factors σ^S, σ^{32}, and σ^{54} compared to the vegetative σ^{70} factor (Fig. 5).[79–81] These negative and positive effects of (p)ppGpp collectively reprogram the cell's transcriptional machinery during transitions between nutritional states, with particularly large effects on the quantity of stable RNAs and ribosomes. In addition to this already complex combination of processes, the RNA polymerase σ^S-holoenzyme ($E\sigma^S$), which transcribes genes involved in resistance to various stresses including starvation,[61] differs in several respects from the σ^{70}-holoenzyme ($E\sigma^{70}$). In particular, $E\sigma^S$ and $E\sigma^{70}$ differ in their sensitivities to (p)ppGpp,[79] DNA supercoiling,[82,83] and intrinsic features of promoters including the spacer region between the −35 and −10 boxes.[84]

During the long-term experiment, most of the lines substituted beneficial mutations in *spoT*. Although these mutations have no measurable effect on the overall rates of synthesis and degradation of (p)ppGpp, we suggest that the evolved cells are poised in stationary phase to respond more quickly to the nutrients that are predictably supplied each day. As we noted, the evolved bacteria recommence growth more quickly than their ancestor after transfer into fresh medium.[20] The increased level of DNA superhelicity in the evolved lines may amplify this phenomenon by facilitating transcription through the discriminator sequence of stringent promoters. Starvation is not a risk, because neither the evolved lines nor the ancestral strain experience any appreciable mortality during the 16–18 h spent each day in stationary phase.[20] We are currently engaged in reconstructing the historical sequence of mutations that were substituted in one of the focal populations, in order to investigate the interactions among the mutations in the precise genetic context in which they arose. In that way, we seek to elucidate the molecular and physiological mechanisms responsible for the improved performance of the bacteria in the environment where they evolved.

Conclusions and synthesis

Evolution experiments allow one to make rigorous connections between genetic changes and phenotypic outcomes in the context of a complex adaptive system, such as a bacterial cell, that includes regulatory networks at multiple levels ranging from operons that involve a few genes to global regulons that may involve hundreds of genes. The mutations discovered by evolution experiments are typically advantageous to the organism that acquires them, at least under the conditions of the experiment, and they are often subtle in their phenotypic effects. In these respects, they are usually different from those mutations found by more traditional genetic studies, in which genes are often knocked out and selective screens usually rely on extreme "plus-or-minus" phenotypes. In effect, experimental evolution focuses attention on the single most-important, integrative and complex phenotype of all, which is the fitness of an organism in a particular environment.

Our review of the long-term evolution experiment with *E. coli* has emphasized two striking outcomes. The first of these is the widespread parallelism, observed at various levels, across the replicate populations even after tens of thousands of generations of independent evolution. At the phenotypic level, all twelve populations have substantially increased fitness, and all of them have cells that are much larger than those of the common ancestor. All the populations also exhibit increased specialization on glucose with associated reductions in other catabolic functions. Most of the populations also show parallel changes in DNA supercoiling, and two focal populations that were studied in greater detail have evolved similar changes in global gene expression profiles at both the transcriptional and proteomic levels.

These parallel phenotypic changes overlie a high level of genetic parallelism, with beneficial mutations found in the same genes in many or all of the evolving populations. By contrast, when randomly chosen genes were sequenced, almost no mutations were found; the few mutations discovered in these random genes occurred only in the subset of populations that had evolved mutator phenotypes, and they showed no parallelism between populations.

Although the same target genes changed repeatedly, the beneficial mutations were different at the nucleotide level in the various populations. The fact that the same genes evolved, whereas the precise mutations were different, raises the question of whether the mutations exerted their beneficial effects via the same or different molecular mechanisms. This issue is of particular interest because many of the target genes are regulatory proteins, such that mutations in those genes may have heterogeneous and widespread pleiotropic effects. Detailed analyses of two regulatory genes, *malT*[38] and *spoT* (our unpublished data), support the conclusion that the different mutations in each of these genes exert their beneficial effects via the same or similar mechanisms.

The second striking outcome seen in this long-term experiment is the prevalence of fitness-enhancing changes in global regulatory networks. Although the precise mechanisms involved in these changes require further investigation, it is clear that evolution of *E. coli* under these conditions—with sudden and predictable transitions between feast and famine—has reshaped and improved the interplay between the DNA superhelicity and stringent response networks. These networks presumably evolved in nature to allow cells to acclimate physiologically to changing conditions, and

thereby avoid the costs (in terms of death and delayed reproduction) of re-evolving the appropriate response each time their environment changed. The changes in the structure and physiological behavior of these regulatory networks that evolved during the long-term experiment clearly demonstrate that they can become better tuned to any particular set of environmental transitions. To paraphrase the philosopher George Santayana: ''Those bacteria who do not learn from history are doomed to repeat it.''

Acknowledgments

We thank the other members of our research groups and outside collaborators for many contributions to this work, including especially Mike Cashel and Tim Cooper.

NP acknowledges the ministry of Education Nationale, de l'Enseignement Supérieur et de la Recherche for a research fellowship and the Université Joseph Fourier for ATER funding. EC acknowledges the Centre National de la Recherche Scientifique for a BDI fellowship.

References

1. Darwin C. 1859. On the origin of species by means of natural selection Murray, London.
2. Luria SE, Delbrück M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511.
3. Elena SF, Lenski RE. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nature Rev Genet 4:457–469.
4. McAdams HH, Srinivasan B, Arkin AP. 2004. The evolution of genetic regulatory systems in bacteria. Nature Rev Genet 5:1–9.
5. Hatfield GW, Benham CJ. 2002. DNA topology-mediated control of global gene expression in *Escherichia coli*. Annu Rev Genet 36:175–203.
6. The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87.
7. Wilkins AS. 2002. Evolution of developmental pathways Sunderland MA: Sinauer.
8. Cashel M, Gentry VJ, Hernandez VJ, Vinella D. 1996. The stringent response. In Neidhardt FC, Curtiss R III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE, editors. *Escherichia coli* and *Salmonella*: cellular and molecular biology Washington DC: ASM Press. p 1458–1496.
9. Travers A, Muskhelishvili G. 2005. DNA supercoiling—a global transcriptional regulator for enterobacterial growth? Nature Rev Microbiol 3:157–169.
10. Errington J. 2003. Regulation of endospore formation in *Bacillus subtilis*. Nature Rev Microbiol 1:117–126.
11. Flärdh K. 2003. Growth polarity and cell division in *Streptomyces*. Curr Opin Microbiol 6:564–571.
12. Kaiser D. 2004. Signaling in myxobacteria. Annu Rev Microbiol 58: 75–98.
13. Huisman GW, Siegele DA, Zambrano MM, Kolter R. 1996. Morphological and physiological changes during stationary phase. In Neidhardt FC, Curtiss R III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE, editors. *Escherichia coli* and *Salmonella*: cellular and molecular biology. Washington DC: ASM Press. p 1672–1682.
14. Moulder JW. 1991. Interaction of *Chlamydiae* and host cells in vitro. Microbiol Rev 55:143–190.
15. Swanson MS, Hammer BK. 2000. *Legionella pneumophila* pathogenesis: a fateful journey from amoebae to macrophages. Annu Rev Microbiol 54: 567–613.
16. Lenski RE. 2004. Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. Plant Breed Rev 24:225–265.
17. 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.
18. Lenski RE, Travisano M. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proc Natl Acad Sci USA 91:6808–6814.
19. Cooper VS, Lenski RE. 2000. The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature 407: 736–739.
20. Vasi F, Travisano M, Lenski RE. 1994. Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. Am Nat 144:432–456.
21. Novak M, Pfeiffer T, Lenski RE, Sauer U, Bonhoeffer S. 2006. Experimental tests for an evolutionary trade-off between growth rate and yield in *E. coli*. Am Nat 168:242–251.
22. Lenski RE, Mongold J. 2000. Cell size, shape, and fitness in evolving populations of bacteria. In Brown JH, West GB, editors. Scaling in biology Oxford: Oxford University Press. p 221–235.
23. Cooper VS, Schneider D, Blot M, Lenski RE. 2001. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J Bacteriol 183:2834–2841.
24. Dekel E, Alon U. 2005. Optimality and evolutionary tuning of the expression level of a protein. Nature 436:588–592.
25. Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. Genetics 148:1667–1686.
26. Sniegowski PD, Gerrish PJ, Lenski RE. 1997. Evolution of high mutation rates in experimental populations of *Escherichia coli*. Nature 387: 703–705.
27. De Visser JAGM, Zeyl CW, Gerrish PJ, Blanchard JL, Lenski RE. 1999. Diminishing returns from mutation supply rate in asexual populations. Science 283:404–406.
28. Shaver AC, Sniegowski PD. 2002. Spontaneously arising *mutL* mutators in evolving *Escherichia coli* populations are the result of changes in repeat length. J Bacteriol 185:6076–6082.
29. Lenski RE, Winkworth CL, Riley MA. 2003. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. J Mol Evol 56:498–508.
30. Chandler M, Mahillon J. 2002. Insertion sequences revisited. In Craig NL, Craigie R, Gellert M, Lambowitz AM. editors. Mobile DNA II. Washington DC: American Society of Biology. p 305–366.
31. 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.
32. Schneider D, Lenski RE. 2004. Dynamics of insertion sequence elements during experimental evolution of bacteria. Res Microbiol 155:319–327.
33. 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.
34. Woods R, Schneider D, Winkworth C, Riley MA, Lenski RE. 2006. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. Proc Natl Acad Sci USA 103:9107–9112.
35. Simpson GG. 1953. The major features of evolution New York: Columbia University Press.
36. Cunningham CW, Jeng K, Husti J, Badgett M, Molineux IJ, Hillis DM, Bull JJ. 1997. Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. Mol Biol Evol 14:113–116.
37. Cooper TF, Rozen DE, Lenski RE. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proc Natl Acad Sci USA 100:1072–1077.
38. Pelosi L, Kühn L, Guetta D, Garin J, Geiselmann J, Lenski RE, Schneider D. 2006. Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. Genetics 173:1851–1869.

39. Crozat E, Philippe N, Lenski RE, Geiselmann J, Schneider D. 2005. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. Genetics 169:523–532.
40. Bledig SA, Ramseier TM, Saier MH Jr. 1996. FruR mediates catabolite activation of pyruvate kinase (*pykF*) gene expression in *Escherichia coli*. J Bacteriol 178:280–283.
41. Begg KJ, Donachie WD. 1998. Division planes alternate in spherical cells of *Escherichia coli*. J Bacteriol 180:2564–2567.
42. Penfound T, Foster JW. 1999. NAD-dependent DNA-binding activity of the bifunctional NadR regulator of *Salmonella typhimurium*. J Bacteriol 181:648–655.
43. Grose JH, Bergthorsson U, Roth JR. 2005. Regulation of NAD synthesis by the trifunctional NadR protein of *Salmonella enterica*. J Bacteriol 187: 2774–2782.
44. Pedersen K, Gerdes K. 1999. Multiple *hok* genes on the chromosome of *Escherichia coli*. Mol Microbiol 32:1090–1102.
45. Reyes-Dominguez Y, Contreras-Ferrat G, Ramirez-Santos J, Membrillo-Hernandez J, Gomez-Eichelmann MC. 2003. Plasmid DNA supercoiling and gyrase activity in *Escherichia coli* wild-type and *rpoS* stationary-phase cells. J Bacteriol 185:1097–1100.
46. Champoux JJ. 2001. DNA topoisomerases: structure, function, and mechanism. Annu Rev Biochem 70:369–413.
47. Dorman CJ, Deighan P. 2003. Regulation of gene expression by histone-like proteins in bacteria. Curr Opinion Genet Dev 13:179–184.
48. Wang JC. 1971. Interaction between DNA and an *Escherichia coli* protein omega. J Mol Biol 55:523–533.
49. Drolet M. 2006. Growth inhibition mediated by excessive negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. Mol Microbiol 59:723–730.
50. Mallik P, Pratt TS, Beach MB, Bradley MD, Undamatla J, Osuna R. 2004. Growth-phase dependent regulation and stringent control of *fis* are conserved processes in enteric bacteria and involve a single promoter (*fis* P) in *Escherichia coli*. J Bacteriol 186:122–135.
51. Weinstein-Fischer D, Elgrably-Weiss M, Altuvia S. 2000. *Escherichia coli* response to hydrogen peroxide: a role for DNA supercoiling, Topoisomerase I and Fis. Mol Microbiol 35:1413–1420.
52. Weinstein-Fischer D, Altuvia S. 2007. Differential regulation of *Escherichia coli* topoisomerase I by Fis. Mol Microbiol 63:1131–1144.
53. Schneider R, Travers A, Kutateladze T, Muskhelishvili G. 1999. A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. Mol Microbiol 34:953–964.
54. Gentry DR, Cashel M. 1996. Mutational analysis of the *Escherichia coli spoT* gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation. Mol Microbiol 19:1373–1384.
55. Mechold U, Murphy H, Brown L, Cashel M. 2002. Intramolecular regulation of the opposing (p)ppGpp catalytic activities of Rel_{Seq}, the Rel/Spo enzyme from *Streptococcus equisimilis*. J Bacteriol 184:2878–2888.
56. Danot O. 2001. A complex signalling module governs the activity of MalT, the prototype of an emerging transactivator family. Proc Natl Acad Sci USA 98:435–440.
57. Cashel M. 1969. The control of ribonucleic acid in *Escherichia coli*. IV. Relevance of unusual phosphorylated compounds from amino acid starved stringent strains. J Biol Chem 244:3133–3141.
58. Cashel M, Gallant J. 1969. Two compounds implicated in the function of the RC gene of *Escherichia coli*. Nature 221:838–841.
59. Lazzarini RA, Dahlberg AE. 1971. The control of ribonucleic acid synthesis during amino acid deprivation in *Escherichia coli*. J Biol Chem 246:420–429.
60. Magnusson LU, Farewell A, Nyström T. 2005. ppGpp: a global regulator in *Escherichia coli*. Trends Microbiol 13:236–242.
61. Hengge-Aronis R. 2002. Signal transduction and regulatory mechanisms involved in control of the σ^S (RpoS) subunit of RNA polymerase. Microbiol Mol Biol Rev 66:373–395.
62. Wendrich TW, Blaha G, Wilson DN, Marahiel MA, Nierhaus KH. 2002. Dissection of the mechanism for the stringent factor RelA. Mol Cell 10: 779–788.
63. Cashel M. 1994. Detection of (p)ppGpp accumulation patterns in *Escherichia coli* mutants. In Adolph KW, editor. Methods in molecular genetics, vol 3, part A. New York NY: Academic Press. p 341–356.
64. Edwards RJ, Sockett RE, Brookfield JF. 2002. A simple method for genome-wide screening for advantageous insertions of mobile DNAs in *Escherichia coli*. Curr Biol 12:863–867.
65. Butland G, Peregrin-Alvarez JM, Li J, Yang W, Yang X, et al. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature 433:531–537.
66. Wout P, Pu K, Sullivan SM, Reese V, Zhou S, Lin B, Maddock JR. 2004. The *Escherichia coli* GTPase $CgtA_E$ cofractionates with the 50S ribosomal subunit and interacts with SpoT, a ppGpp synthetase/hydrolase. J Bacteriol 186:5249–5257.
67. Battesti A, Bouveret E. 2006. Acyl carrier protein/SpoT interaction, the switch linking SpoT-dependent stress response to fatty acid metabolism. Mol Microbiol 62:1048–1063.
68. Barabasi A-L. Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. Nature Rev Genet 5:101–113.
69. Koch AL. 1971. The adaptive responses of *Escherichia coli* to a feast and famine existence. Adv Microb Physiol 6:147–217.
70. Bremer H, Dennis PP. 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In Neidhardt FC, Curtiss R III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE, editors. *Escherichia coli* and *Salmonella*: cellular and molecular biology Washington DC ASM Press: p 1553–1569.
71. Travers A. 1976. Modulation of RNA polymerase specificity by ppGpp. Mol Gen Genet 147:225–232.
72. Lamond AI. 1985. Supercoiling response of a bacterial tRNA gene. EMBO J 4:501–507.
73. Muskhelishvili G, Travers AA, Heumann H, Kahmann R. 1995. FIS and RNA polymerase holoenzyme form a specific nucleoprotein complex at a stable RNA promoter. EMBO J 14:1446–1452.
74. Zhou YN, Jin DJ. 1998. The *rpoB* mutants destabilizing initiation complexes at stringently controlled promoters behave like ''stringent'' RNA polymerases in *Escherichia coli*. Proc Natl Acad Sci USA 95:2908–2913.
75. Travers AA. 1980. Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. J Bacteriol 141:973–976.
76. Pemberton IK, Muskhelishvili G, Travers AA, Buckle M. 2000. The GC-rich discriminator region of the *tyrT* promoter antagonises the formation of stable pre-initiation complexes. J Mol Biol 299:859–864.
77. Barker MM, Gaal T, Gourse RL. 2001. Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. J Mol Biol 305: 689–702.
78. Paul BJ, Berkmen MB, Gourse RL. 2005. DksA potentiates direct activation of amino acid promoters by ppGpp. Proc Natl Acad Sci USA 102:7823–7828.
79. Jishage M, Kvint K, Shingler V, Nyström T. 2002. Regulation of sigma factor competition by the alarmone ppGpp. Genes Dev 16:1260–1270.
80. Kvint K, Farewell A, Nyström T. 2000. RpoS-dependent promoters require guanosine tetraphosphate for induction even in the presence of high levels of sigma(S). J Biol Chem 275:14795–14798.
81. Laurie AD, Bernardo LM, Sze CC, Skarfstad E, Szalewska-Palasz, A et al.The role of the alarmone (p)ppGpp in sigma N competition for core RNA polymerase. J Biol Chem 278:1494–1503.
82. Kusano S, Ding Q, Fujita N, Ishihama A. 1996. Promoter selectivity of *Escherichia coli* RNA polymerase $E\sigma^{70}$ and $E\sigma^{38}$ holoenzymes. J Biol Chem 271:1998–2004.
83. Bordes P, Conter A, Morales V, Bouvier J, Kolb A, Gutierrez C. 2003. DNA supercoiling contributes to disconnect σ^S accumulation from σ^S-dependent transcription in *Escherichia coli*. Mol Microbiol 48:561–571.
84. Typas A, Hengge R. 2006. Role of the spacer between the -35 and -10 regions in σ^S promoter selectivity in *Escherichia coli*. Mol Microbiol 59: 1037–1051.