Parallel Genetic and Phenotypic Evolution of DNA Superhelicity in Experimental Populations of Escherichia coli

Estelle Crozat,1,2 Cynthia Winkworth,†,3 Joël Gaffé,1,2 Peter F. Hallin,4 Margaret A. Riley,3 Richard E. Lenski,5 and Dominique Schneider*,1,2

1Laboratoire Adaptation et Pathogénie des Micro-Organismes, Université Joseph Fourier Grenoble 1, Grenoble cedex 9, France
2CNRS UMR 5163, Laboratoire Adaptation et Pathogénie des Microorganismes, Grenoble, France
3Department of Biology, University of Massachusetts
4Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark
5Department of Microbiology and Molecular Genetics, Michigan State University

†Present address: Department of Zoology, University of Otago, Dunedin, New Zealand.

*Corresponding author: E-mail: dominique.schneider@ujf-grenoble.fr.
Associate editor: Douglas Crawford

Abstract

DNA supercoiling is the master function that interconnects chromosome structure and global gene transcription. This function has recently been shown to be under strong selection in Escherichia coli. During the evolution of 12 initially identical populations propagated in a defined environment for 20,000 generations, parallel increases in DNA supercoiling were observed in ten populations. The genetic changes associated with the increased supercoiling were examined in one population, and beneficial mutations in the genes topA (encoding topoisomerase I) and fis (encoding a histone-like protein) were identified. To elucidate the molecular basis and impact of these changes, we quantified the level of genetic, phenotypic, and molecular parallelism linked to DNA supercoiling in all 12 evolving populations. First, sequence determination of DNA topology-related loci revealed strong genetic parallelism, with mutations concentrated in three genes (topA, fis, and dusB), although the populations had different alleles at each locus. Statistical analyses of these polymorphisms implied the action of positive selection and, moreover, suggested that fis and dusB, which belong to the same operon, have related functions. Indeed, we demonstrated that DusB regulates the expression of fis by both experimental and phylogenetic analyses. Second, molecular analyses of five mutations in fis and dusB affecting the transcription, translation, and protein activity of Fis also revealed strong parallelism in the resulting phenotypic effects. Third, artificially increasing DNA supercoiling in one of the two populations that lacked DNA topology changes led to a significant fitness increase. The high levels of molecular and genetic parallelism, targeting a small subset of the many genes involved in DNA supercoiling, indicate that changes in DNA superhelicity have been important in the evolution of these populations. Surprisingly, however, most of the evolved alleles we tested had either no detectable or slightly deleterious effects on fitness, despite these signatures of positive selection.

Key words: evolution, bacteria, DNA superhelicity, fitness, phenotypic and genetic parallelism.

Introduction

DNA superhelicity controls one of the most important regulatory networks in bacteria (Travers and Muskhelishvili 2005). It is involved in such vital processes as chromosome replication and segregation, transcription, translation, recombination, and cellular responses to environmental conditions. Many external stimuli produce transient modifications of DNA superhelicity, including changes in nutritional conditions (Balke and Gralla 1987; Reyes-Dominguez et al. 2003), oxidative stress (Weinstein-Fischer et al. 2000), osmotic stress (Higgins et al. 1988), thermal stress (Goldstein and Drlica 1984), pH stress (Karem and Foster 1993), presence or absence of oxygen (Bhriain et al. 1989), and intracellular growth of pathogens (O’Croinin et al. 2006). These transient alterations of superhelicity in turn modify the cell’s global transcriptional patterns, thereby allowing phenotypic acclimation of bacteria to their environment (Higgins et al. 1988; Weinstein-Fischer et al. 2000; Peter et al. 2004).

DNA superhelicity impacts transcription through two mechanisms. The first acts locally, whereby the level of supercoiling alters the physical properties of promoter sequences, for example, by influencing the opening of the DNA duplex (Dai and Rothman-Denes 1999). The second mechanism acts more globally and involves the organization of the bacterial chromosome into topological domains (Postow et al. 2004; Valens et al. 2004; Deng et al. 2005). This domain organization is correlated with the existence of spatial patterns of transcription in the Escherichia coli genome (Jeong et al. 2004), which can be altered by modulating superhelicity. Structural organization of the chromosome, DNA superhelicity, and global transcription patterns are therefore highly interconnected processes that allow cells to cope with fluctuating environmental conditions.

The level of DNA superhelicity is tightly controlled by the combined activities of topoisomerases (Champoux 2001) and histone-like proteins (Dorman and Deighan...
Virulence genes (O’Croinin et al. 2006). (Lenski et al. 1991; Cooper and Lenski 2000; Lenski 2001) in a minimal medium supplemented with limiting glucose more than 40,000 generations by daily serial transfer in the longest running evolution experiment, Lenski and colleagues have propagated 12 populations of E. coli for more than 50,000 generations, which about one-quarter had been identified by the previous study did not investigate whether parallel mutations had produced the parallel increases in DNA superhelicity observed in the other populations. Several other phenotypic traits evolved in parallel in most or all populations, including cell size (Lenski et al. 1998; Philippe et al. 2009), growth parameters (Vasi et al. 1994), catabolic functions (Cooper and Lenski 2000; Pelosi et al. 2006), and global gene expression profiles. Genetic studies aimed at understanding these parallel phenotypic changes identified several genes that were also reproducibly altered in most or all populations, including rbs, the ribose utilization operon (Cooper et al. 2001); spoT, a gene involved in (p)ppGpp metabolism during the stringent response (Cooper et al. 2003; Pelosi et al. 2006); and malT, the transcriptional activator of the maltose utilization operons (Pelosi et al. 2006). Other parallel genetic changes were discovered in nadR, involved in NAD metabolism; pykF, which encodes pyruvate kinase I; ppbA-rodA, involved in cell wall biosynthesis; and hokB-sockB, a plasmid maintenance module (Schneider, Duperchy et al. 2000; Woods et al. 2006). Many of these genetic changes were demonstrated, by constructing and competing otherwise isogenic strains, to be beneficial in the environment of the evolution experiment. Complete genome sequences were recently obtained for evolved clones isolated at several generations from one population. A total of 45 mutations were found in the clone sampled after 20,000 generations (Barrick et al. 2009), of which about one-quarter had been identified by the previous genetic studies.

DNA supercoiling provides an interesting candidate function for further investigation into the extent as well as the pattern of evolutionary parallelism in these populations because more than 20 genes (encoding topoisomerases, histone-like proteins, and gene expression regulators) are involved in its regulation in E. coli. For the present study, we analyzed the extent of parallelism and its underlying molecular basis in three ways. First, we sequenced nine topology-related loci (including topoA and fis) in the ancestor and evolved clones from all 12 populations. Under one hypothesis, mutations in many different genes that control DNA superhelicity may have caused the parallel increases in supercoiling. Alternatively, the same one or few genes may have been targeted by selection.
Table 1. The 17 Topology-Related Genes and Upstream Open Reading Frames Sequenced in this Study.

<table>
<thead>
<tr>
<th>Gene (upstream Open Reading Frames)</th>
<th>Number of Base Pairs Sequenced</th>
<th>Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>dps</td>
<td>802</td>
<td>Histone-like protein</td>
<td>Almiron et al. (1992)</td>
</tr>
<tr>
<td>fs (dusB)</td>
<td>1,695</td>
<td>Histone-like protein</td>
<td>Finkel and Johnson (1992)</td>
</tr>
<tr>
<td>gyrA</td>
<td>2,774</td>
<td>Gyrase subunit A</td>
<td>Champoux (2001)</td>
</tr>
<tr>
<td>gyrB</td>
<td>2,673</td>
<td>Gyrase subunit B</td>
<td>Champoux (2001)</td>
</tr>
<tr>
<td>hns</td>
<td>1,160</td>
<td>Histone-like protein</td>
<td>Yamada et al. (1991)</td>
</tr>
<tr>
<td>parC (yqiS)</td>
<td>2,599</td>
<td>Topoisomerase IV subunit A</td>
<td>Peng and Marians (1993)</td>
</tr>
<tr>
<td>parE (yqiA, icc, yqiB, yqiE)</td>
<td>4,652</td>
<td>Topoisomerase IV subunit B</td>
<td>Peng and Marians (1993)</td>
</tr>
<tr>
<td>topA</td>
<td>2,905</td>
<td>Topoisomerase I</td>
<td>Wang (1971)</td>
</tr>
<tr>
<td>topB (selD, ydjA)</td>
<td>3,842</td>
<td>Topoisomerase III</td>
<td>Champoux (2001)</td>
</tr>
</tbody>
</table>

in all of the independently evolving populations. Second, given that substantial genetic parallelism was observed, we analyzed how these changes related to gene regulation and to fitness. Third, two populations showed no increases in their DNA superhelicity (Crozat et al. 2005). We determined whether moving topology-related mutations from another evolved population affected superhelicity and fitness in one of these two populations that retained the ancestral superhelicity.

Materials and Methods

Long-Term Evolution Experiment
The long-term evolution experiment is described in detail elsewhere (Lenski 2004). In brief, 12 populations were founded using two E. coli B clones as ancestors (Jeong et al. 2009) and then propagated by serial daily transfers at 37 °C for 20,000 generations (3,000 days) in Davis minimal medium supplemented with glucose at 25 mg/ml (DM25) (Lenski et al. 1991). Each day, after 24-h incubation, cells were diluted 100-fold into fresh medium. One ancestral strain, REL606, cannot grow on arabinose (Ara−), and it was used to initiate six populations, Ara−1 to Ara−6. The other six populations, Ara+E to Ara+E+, were started with REL607, a spontaneous Ara− mutant of the source strain. The Ara phenotype serves as a neutral marker in competition experiments for measuring relative fitness (Lenski et al. 1991; Lenski 2004).

Strains, Plasmids, and Culture Conditions
Evolved clones used in this study were isolated either at generation 20,000 (for sequence analyses and Western blots) or at generations 2,000 and 5,000 (for allele exchange experiments) from the 12 populations (Cooper and Lenski 2000); strains were stored as 15% glycerol suspensions at −80 °C. Strains that are isogenic except for topA and fis alleles were previously constructed in the ancestral background (Crozat et al. 2005) using the suicide plasmid pKO3 (Link et al. 1997), for allele replacements, and pCRII-Topo (Invitrogen) for cloning experiments. Plasmids pKG13 and pKK223-3 (Gosink et al. 1996) were used to evaluate the DNA-binding activity of ancestral and evolved Fis proteins.

All experiments were performed by growing strains either in the DM25 minimal medium that was used in the evolution experiment (Lenski et al. 1991) or in Luria Broth (LB) (Sambrook et al. 1989). Ampicillin (100 µg/ml), chloramphenicol (30 µg/ml), or kanamycin (50 µg/ml) was added as needed.

DNA Sequencing
Nine loci encoding topoisomerases or histone-like proteins were sequenced in an evolved clone isolated at 20,000 generations from each of the 12 populations. That generation was chosen because most of the previous genetic studies, including the one on DNA topology (Crozat et al. 2005), focused on this time point. Moreover, the focal population Ara−1 evolved a mutator phenotype after 26,000 generations (Barrick et al. 2009), which complicates the analyses. Four of the sequenced genes are located within operons and the whole operons were then sequenced, such that 17 genes in all were sequenced in each clone. Table 1 details the genes sequenced in this study, as well as function of the gene products and number of base pairs sequenced. All 17 genes (totaling 23,102 bp) were sequenced in 14 clones, including the 12 evolved clones as well as the 2 ancestral variants, REL606 and REL607; as expected, the ancestors were identical. Genomic DNA preparations, polymerase chain reaction (PCR) amplifications, sequencing reactions, and sequence analyses were performed as previously described (Lenski et al. 2003). All primer pairs used for the genes in Table 1 are available upon request to the corresponding author. All polymorphic sites were confirmed by resequencing the relevant regions twice with independent DNA extractions and reanalyzing the sequence chromatogram using Sequencher 3.1.1. Polymorphisms were analyzed using the Molecular Evolutionary Analysis Package MEA (Etsuko Moriymaya, Yale University version 6/22/2000), based on the algorithm of Nei and Gojobori (1986).

Construction of Isogenic Strains
Evolved fis alleles from two populations (Ara+E and Ara+E−) and an evolved dusB allele from another (Ara+E+) were moved into the ancestral chromosome by homologous recombination using the suicide plasmid pKO3 (Link et al.
Crozat et al. · MBE

1997), as described previously (Crozat et al. 2005). The constructed strains were checked for the presence of the evolved allele. Both evolved fis alleles generated new restriction sites, and screening employed a PCR–restriction fragment length polymorphism (PCR-RFLP) approach using Alul and Ddel (Euromedex) to distinguish between the ancestral and evolved alleles. To distinguish between ancestral and evolved alleles of dusB, the region of dusB bearing the evolved mutation was sequenced. All constructed strains were also deconstructed, such that the evolved alleles were replaced by their ancestral counterparts, to confirm the absence of secondary mutations during the construction process. These experiments generated the following isogenic strains: 606^fis+^, 606^fis−^, and 606^dusB+^, 606^dusB−^.

All have the ancestral REL606 genome except they carry the indicated evolved alleles. We could not move the evolved fis allele from Ara−4 into the ancestral strain, despite numerous attempts, for reasons unknown. The same strategy was also used to replace the Ara+5 evolved dusB allele by its ancestral allele in a 5,000-generation clone, with the resulting strain called 5K^dusBanc^.

The same strategy was used to move the evolved topA allele from population Ara−1 (Crozat et al. 2005) into the chromosome of an evolved clone, called 1158C, isolated from population Ara+1 at 2,000 generations. The 1158C clone contains the ancestral topA allele, which was thus replaced by the evolved Ara−1 topA allele to produce strain 1158C^topA−^.

**Fitness Assays**

The relative fitness of two strains was estimated by competing them, as described elsewhere (Pelosi et al. 2006). Briefly, isogenic strains were separately competed against the ancestor carrying the opposite arabinose marker. Each pairwise competition was replicated at least 6-fold. Competitions were performed in DM25, the same medium used for the evolution experiment, for either 1 or 6 days, with the longer assays used to measure smaller fitness differences. Samples were taken immediately after mixing at day 0 and again after 1 or 6 days of competition to measure the abundance of both competitors. Using the initial and final cell counts and dilution factors allows one to calculate the realized (net) population growth of each competitor. Fitness was then calculated as the ratio of their realized growth rates, and t-tests were used to evaluate whether the values differed significantly from the null hypothetical value of one. To determine the fitness effect of the evolved Ara+5 dusB allele in the genetic background of a 5,000-generation clone, the 5K and 5K^dusBanc^ clones competed separately against the REL606 ancestor and their fitness values were compared. We also tried to compete the 5K and 5K^dusBanc^ clones directly against each other. To distinguish them, we introduced the araA mutation responsible for the Ara− phenotype of REL606 (Jeong et al. 2009) into each clone. However, we obtained inconsistent results with these marked pairs of clones, suggesting that a secondary mutation occurred during one of the constructions. In order to achieve better resolution of their relative fitness, we also competed the 5K and 5K^dusBanc^ clones separately against a 5,000-generation clone sampled from population Ara−1.

**DNA Superhelicity Measurements**

Topological changes were measured by the relative abundance of different topoisomers of a reporter plasmid, as described previously (Crozat et al. 2005). Briefly, plasmid pUC18 was introduced by electrottransformation into the strains of interest. After growing cells in LB-ampicillin media to an OD^600 nm^ of 2, plasmids were extracted (Qiagen Qiaprep Spin Miniprep Kit) and quantified, with 200 ng loaded onto 1% agarose gels containing 1.5 μg/ml chloroquine. Levels of DNA supercoiling were estimated by mean σ values. The limitations and validity of this plasmid-based assay are discussed in Crozat et al. (2005).

**Electrophoresis and Immunoblot Analysis of Proteins**

Triplicate cultures of the ancestral strains bearing the different fis and dusB alleles, and of evolved clones sampled at 20,000 generations from the 12 populations, were grown in LB medium at 37 °C. Aliquots were sampled during exponential phase after 1.5 h of incubation. After centrifugation, cells were resuspended in lysis buffer (70 mM Tris pH 7.4, 1 mM ethylenediaminetetraacetic acid [EDTA], 1 mM dithiothreitol [DTT], and 10% glycerol) before sonication. Lysates were centrifugated at 10,000 × g for 30 min at 4 °C, and total protein concentrations were determined using the Bradford Protein Assay Kit (Bio-Rad) with bovine serum albumin as a standard. Equal amounts of protein were loaded onto 16% sodium dodecyl sulfate–polyacrylamide gels, separated by electrophoresis and electrotransferred (Bio-Rad) onto nitrocellulose membranes (Amersham Pharmacia). Prestained protein standards (Euromedex) were used to estimate molecular weights. Immunoblot analyses were performed with antibodies against Fis (courtesy of A. Ishihama, Nippon Institute for Biological Sciences, Japan) and RpoA (courtesy of M. Cashel, National Institutes of Health, USA) and developed using the ECL plus Western Blotting Detection System Kit (Amersham Pharmacia).

Cell cultures for these assays were grown in LB rather than DM25 for three reasons. First, LB gave higher cell densities than DM25, especially during early exponential phase. Second, we wanted to be consistent with previous analyses (Crozat et al. 2005). The previous work, moreover, had demonstrated that the observed changes in superhelicity in the evolved clones were consistent between LB and DM25 media, although it was much more difficult to make the measurements in DM25 (Crozat et al. 2005). Third, it was previously shown that the expression profile of Fis, with a peak in exponential phase, was similar in LB and minimal media (Ball et al. 1992; Mallik et al. 2004), although the maximum level is lower in minimal medium.

**Transcriptional Fusions and β-Galactosidase Assays**

Ancestral and evolved Ara+5 fis promoter alleles were PCR amplified from positions −241 to +84 relative to the +1 dusB transcription start site (Mallik et al. 2004) using
The OD600 nm of 0.6–0.8, and Fis expression was induced with IPTG. The expression vector and overexpressed in BL21(DE3) cells. The fisBamHI site was introduced into the fis derivatives were then introduced into the chromosome of ancestral strain REL606. The 5′-att site in strains 606Δatt5 and 606Δatt6. For both fusions, the sequences of the fis promoters were verified, and the single lysogenization of fusion-carrying λ phages was checked by using a previously described PCR method (Powell et al. 1994).

For β-galactosidase assays, the cultures were grown in kanamycin-containing LB media, and samples were taken at different times during growth. The β-galactosidase specific activities were assayed by using o-nitrophenyl β-D-galactopyranoside (ONPG) as a substrate, and they were expressed in nanokatal per milligram (nanomoles of β-galactosidase assays, the cultures were grown in 15 μl solution containing 20 mM Tris pH 7.5, 1 mM EDTA pH 8, 80 mM NaCl, 4% glycerol, 0.5 mM DTT, 0.5 mg/ml bovine serum albumin, and 10 μg/ml of sonicated herring sperm DNA. After 30 min of incubation at 37°C, 5 μl of loading buffer (0.08% bromophenol blue in 50% sucrose) was added and samples were electrophoresed on 5% polyacrylamide gels (19:1) in 0.5× Tris-Borate EDTA buffer. The gels were then dried, exposed to FujiFilm Imaging Plates, and revealed on an FLA8000 scanner (Fujifilm). The percentage of bound DNA was quantified with the Image Gauge software (FujiFilm). Several types of controls were performed. From 5 to 1,000-fold excesses of unlabeled DNA fragments were added to the binding mixtures. The fragments consisted of a 30-bp oligonucleotide containing either a consensus Fis-binding site or the same oligonucleotide except with the Fis site mutated. A supershift assay was also performed by adding the antibody raised against Fis to the binding reactions.

Results

Evidence for Selection on Topology-Related Genes

DNA sequences were determined for 17 candidate genes, distributed over nine operons, for the ancestor and for clones sampled from all 12 evolving populations at generation 20,000 (Table 2). These operons are involved in DNA supercoiling regulation in E. coli; six encode DNA topoisomerases and three major histone-like proteins. The 12 evolved clones have a total of 27 mutations.

To evaluate whether any of the candidate genes known to be involved in DNA topology, or any other genes in the same operons, were targets of selection during the long-term evolution experiment, we applied four statistical tests (Woods et al. 2006). The first test compares the overall rate of substitution in these 17 genes versus 36 previously sequenced randomly chosen gene regions (Lenski et al. 2003). A total of 23,102 bp were sequenced for the topology genes, whereas 18,374 bp were sequenced for the random genes. Table 3 shows that 11 of the 12 populations had higher rates of change in the topology-related genes than in the random genes. This difference is highly significant (one-tailed sign test, P = 0.0032) in the direction predicted...
Table 2. Mutations Found in 12 Evolving Populations after 20,000 Generations of Evolution among Topology-Related Genes and Upstream Open Reading Frames.

<table>
<thead>
<tr>
<th>Population</th>
<th>Gene</th>
<th>Nucleotide Position</th>
<th>Mutation</th>
<th>Amino Acid Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ara−1</td>
<td>fs</td>
<td>−4</td>
<td>A To C</td>
<td>NA (RBS)</td>
</tr>
<tr>
<td>Ara−2 (M)</td>
<td>topA</td>
<td>+97</td>
<td>C To T</td>
<td>H33 To Y</td>
</tr>
<tr>
<td></td>
<td>topB</td>
<td>+628</td>
<td>C To T</td>
<td>D210 To N</td>
</tr>
<tr>
<td></td>
<td>ydiA (topB locus)</td>
<td>+270</td>
<td>G To A</td>
<td>No change</td>
</tr>
<tr>
<td></td>
<td>ydiA (topB locus)</td>
<td>+15</td>
<td>T To C</td>
<td>No change</td>
</tr>
<tr>
<td>Ara−3</td>
<td>fis</td>
<td>+152</td>
<td>A To C</td>
<td>Y51 To S</td>
</tr>
<tr>
<td></td>
<td>topA</td>
<td>+2315</td>
<td>C To A</td>
<td>T792 To K</td>
</tr>
<tr>
<td>Ara−4 (M)</td>
<td>fis</td>
<td>+177</td>
<td>+ A</td>
<td>CTD change</td>
</tr>
<tr>
<td>Ara−5</td>
<td>topA</td>
<td>−112</td>
<td>G To T</td>
<td>NA (promoter)</td>
</tr>
<tr>
<td></td>
<td>topA</td>
<td>+599</td>
<td>C To A</td>
<td>A200 To E</td>
</tr>
<tr>
<td>Ara−6</td>
<td>dusB (fis locus)</td>
<td>+904</td>
<td>G To A</td>
<td>A302 To T</td>
</tr>
<tr>
<td>Ara+1</td>
<td>fis</td>
<td>−63</td>
<td>+</td>
<td>NA (promoter)</td>
</tr>
<tr>
<td>Ara+2</td>
<td>fis</td>
<td>+329</td>
<td>C To A</td>
<td>NA (downstream)</td>
</tr>
<tr>
<td>Ara+3 (M)</td>
<td>dusB (fis locus)</td>
<td>+936</td>
<td>G To A</td>
<td>No change</td>
</tr>
<tr>
<td></td>
<td>parC</td>
<td>+460</td>
<td>G To A</td>
<td>P154 To S</td>
</tr>
<tr>
<td>Ara+4</td>
<td>dusB (fis locus)</td>
<td>+942</td>
<td>+ 9 bp</td>
<td>+ ALE after E314</td>
</tr>
<tr>
<td></td>
<td>dps</td>
<td>+143</td>
<td>IS1</td>
<td>NA</td>
</tr>
<tr>
<td>Ara+5</td>
<td>dusB (fis locus)</td>
<td>+943</td>
<td>G To A</td>
<td>A315 To T</td>
</tr>
<tr>
<td>Ara+6 (M)</td>
<td>dusB (fis locus)</td>
<td>+899</td>
<td>T To G</td>
<td>F300 To C</td>
</tr>
<tr>
<td></td>
<td>selD (topB locus)</td>
<td>+960</td>
<td>T To G</td>
<td>No change</td>
</tr>
<tr>
<td></td>
<td>yqiB (parC locus)</td>
<td>+134</td>
<td>T To G</td>
<td>Y45 To S</td>
</tr>
<tr>
<td></td>
<td>yqiA (parC locus)</td>
<td>+251</td>
<td>A To C</td>
<td>L84 To R</td>
</tr>
<tr>
<td></td>
<td>hns</td>
<td>−89</td>
<td>A To C</td>
<td>NA (promoter)</td>
</tr>
<tr>
<td></td>
<td>gyrA</td>
<td>+1925</td>
<td>T To G</td>
<td>Q657 To H</td>
</tr>
<tr>
<td></td>
<td>gyrB</td>
<td>+1971</td>
<td>T To G</td>
<td>E662 To A</td>
</tr>
</tbody>
</table>

Note.—NA, not applicable.

Populations showing a mutator phenotype (Snigovski et al. 1997) are indicated by (M).

Nucleotide positions are given relative to the first position of the translational start codon of the corresponding gene unless indicated otherwise.

Mutation within the ribosome binding site.

Mutation within the promoter region.

The nucleotide position is given relative to the first position of the translational start codon of dusB.

Table 3. Number of Mutations in Random and Topology-Related Genes in 12 Escherichia coli Populations that Evolved for 20,000 Generations.

<table>
<thead>
<tr>
<th>Population</th>
<th>Random Genes (18,374 bp total)</th>
<th>Topology-Related Genes (23,102 bp total)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of Mutations</td>
<td>Rate per 1,000 bp</td>
</tr>
<tr>
<td>Ara−1</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Ara−2 (M)</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Ara−3</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Ara−4 (M)</td>
<td>3</td>
<td>0.163</td>
</tr>
<tr>
<td>Ara−5</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Ara−6</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Ara+1</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Ara+2</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>Ara+3 (M)</td>
<td>1</td>
<td>0.054</td>
</tr>
<tr>
<td>Ara+4</td>
<td>2</td>
<td>0.109</td>
</tr>
</tbody>
</table>

Populations Ara−2, Ara−4, Ara−3, and Ara−6 became mutators (M), whereas the others retained the low ancestral mutation rate.

Including those mutations at or near fixation (Lenski et al. 2003).

Topology-related genes include regulatory regions as well as other ORFs found in the same loci.

if one or more of the topology-related genes were undergoing accelerated evolution, consistent with being targets of selection. The difference is significant even if we exclude population Ara−1 (P = 0.0059), which had previously been shown to harbor beneficial mutations in both topA and fis (Crozat et al. 2005).

Of the 27 mutations identified in the topology-related genes, 11 (40.7%) were in topA, fis, or their regulatory regions, although together these two genes constituted only 3,634 bp (15.7%). Another gene, dusB, had five mutations (18.5%) in only 966 bp of coding sequence (4.2%). The dusB gene encodes one of three dihydrouridine synthases in E. coli (Bishop et al. 2002). Dihydrouridine is the most abundant modified base found at conserved positions in the D-loop of transfer RNA (tRNA) in prokaryotes and eukaryotes, but its physiological role is unknown (Bishop et al. 2002). The other 11 mutations (40.7%) were dispersed across ten different genes that comprised 80.1% of the sequencing effort, including 2 in ydiA and 1 each in topB, parC, dps, selD, yqiB, yqiA, hns, gyrA, and gyrB. We conclude that mutations in dusB, as well as those in topA and fis, provide evidence of positive selection based on this first statistical test. Moreover, the fact that dusB mutations were found in five populations (Ara−6, Ara+3, Ara+4, Ara+5, and Ara+6) that evolved changes in supercoiling (Crozat et al. 2005), while lacking any mutations in topA or fis, suggests that dusB mutations were responsible for the changes in DNA topology. The remaining statistical tests therefore focused on topA, fis, and dusB.

The second test compares the pattern of synonymous and nonsynonymous substitutions in these 3 topology-related genes with the pattern observed in the 36 randomly chosen genes. The random genes had equal numbers of synonymous and nonsynonymous mutations (three each) while nonsynonymous changes predominated in the topology-related genes (11/12 nonsynonymous including insertions in the coding sequences). This difference is, again, in the direction expected if the topology genes experienced selection for changes in the encoded proteins, although the difference is only marginally significant (one-tailed Fisher’s exact test, P = 0.0833).

The third test compares the distribution of point mutations between four populations that became mutators, owing to evolved defects in DNA repair, and eight populations that retained the low ancestral mutation rate (table 3). Mutations in topology genes were more common in the nonmutator (10 mutations) than in the mutator (4 mutations) populations. In contrast, all six mutations in the randomly chosen genes were found in the four mutator populations. This difference is in the expected direction under the hypothesis that both mutator and nonmutator populations accumulate mutations in target genes, whereas mutator populations differentially accumulate mutations in other loci that are not under positive selection (one-tailed
Evolutionary Changes in DNA Superhelicity · doi:10.1093/molbev/msq099

FIG. 1. Substitutions in fis and dusB genes and structure of the dusB-fis operon. The −35 and −10 promoter elements (Malik et al. 2004) are shown by black boxes and coding sequences by white rectangles. The broken arrow marks the initiation site and direction of transcription. The locations of ten mutations substituted in the long-term evolution experiment are indicated by arrows and labeled according to the population in which they arose. In this study, the three evolved fis alleles from populations Ara−1, Ara+1, and Ara−3 and the evolved dusB allele from population Ara+5 were genetically manipulated to construct otherwise isogenic strains. We were unable to move the fis allele from population Ara−4 to the ancestral background despite multiple attempts. However, we purified the evolved Fis protein from Ara−4, together with the corresponding protein from Ara−3, for biochemical analyses. We did not analyze the fis allele from population Ara+2 because it lies 3′ to the dusB-fis operon. All five evolved dusB alleles have mutations within a 45-bp region that encodes the C-terminal region of DusB, which is known not to be involved in the dihydrouridine synthase activity (Savage et al. 2006). We therefore genetically manipulated and studied only one dusB allele from population Ara+5.

The substitution of the evolved topA and fis alleles was previously demonstrated in the Ara−1 population by sequencing those genes in multiple clones sampled at several generations (Crozat et al. 2005). We similarly confirmed fixation of the evolved topA, fis, and dusB alleles in the other populations (table 2). In population Ara−2, the two evolved topA alleles had fixed by generations 10,000 (mutation at position +1526, 3/3 clones tested) and 20,000 (position +1012, 3/3 clones). In Ara−3, the topA allele fixed by generation 4,000 (3/3 clones tested). The two topA mutations in Ara−5 fixed by generations 2,000 (position +599, 3/3 clones tested) and 20,000 (position −112, 3/3 clones). The evolved fus alleles were substituted by generation 7,000 in population Ara−3 (3/3 clones), generation 20,000 in Ara−4 (3/3 clones), generation 20,000 in Ara+1 (3/3 clones), and generation 10,000 in Ara+2 (2/2 clones). The dusB alleles in populations Ara+5 and Ara+6 were substituted by 5,000 generations (13/13 clones) and 10,000 generations (2/2 clones), respectively.

Phenotypic Effects of fis and dusB Mutations Features of Several Evolved Alleles

We observed striking genetic parallelism in terms of the concentration of mutations in three topology-related genes. However, there was also considerable divergence, with different alleles at each locus as well as different combinations of genes with mutations. Therefore, we analyzed the functional effects of several mutations to assess whether they confer similar phenotypes. We focused on four evolved alleles, three with mutations affecting fis and one affecting dusB (table 2). The fis mutations in populations Ara+1, Ara−1, and Ara−3 impact different aspects of fis gene expression (fig. 1). Specifically, the fis allele in Ara−1 has an A-to-C conversion located in the RBS that was previously described (Crozat et al. 2005), which caused a 3-fold decrease in the level of Fis, presumably by affecting translation initiation. The fis allele in Ara+1 presumably affects transcription because it altered the sixth position of the fis promoter −35 element (Walker et al. 1999), a T-to-G transversion at position −31 relative to the +1 transcription start site. This change does not bring the −35 element any closer to the consensus

Fisher’s exact test, $P = 0.0054$). Importantly, and consistent with our separation of topA, fis, and dusB from the other genes in topology-related loci, all ten point mutations in those other genes were in the four mutator lines (one mutation in these other genes was in a nonmutator population, but it was an IS1 insertion in dps). This distributional difference between topA, fis, and dusB and the other topology-related genes is also highly significant (one-tailed Fisher’s exact test, $P = 0.0005$). The other topology-related genes thus accumulate mutations in a manner similar to the set of randomly chosen genes.

The fourth sequence-based statistical test evaluates the distribution of mutations across the 12 independently evolved populations. If mutations accumulated randomly by drift, and if the populations have the same mutation rate, then one would expect a Poisson distribution of these mutations. Given the variation in mutation rates caused by the defects in DNA repair in four populations, one would expect a highly clustered distribution under the drift hypothesis. By contrast, if selection drives sequence evolution and if the different mutations at a given locus, or set of loci, produce the same or similar benefit (such that multiple mutations exhibit diminishing returns), then one would expect a more uniform distribution of mutations across the replicate lines. Two of the candidate genes, fis and dusB, show maximally uniform distributions, each gene harboring single mutations in five different populations, although the number of mutations for either gene alone is too small for statistical significance. However, if we combine these two genes, which together comprise the full extent of an operon (Malik et al. 2004), we see that ten populations each have exactly one mutation in either fis or dusB. That distribution is maximally uniform and significantly different from a random Poisson distribution ($P = 0.0039$). This outcome is driven not only by the uniform distribution of mutations in each separate gene but also by a significant negative association between the two genes (two-tailed Fisher’s exact test, $P = 0.0079$). This pattern therefore led us to analyze the molecular and phenotypic interactions of fis and dusB and their associated gene products. In the next section, we confirm the sequence-based hypothesis that mutations in the fis and dusB genes can produce similar phenotypic effects.
5′-TTGACA-3′ sequence. The Ara−3 fis allele has a Y-to-S amino acid replacement at residue 51, which may affect the activity of the Fis protein. The ancestral Y51 residue is highly conserved, located at the start of the B helix of a Fis monomer, and is involved in the formation of the active Fis dimers (Yuan et al. 1991). Finally, the evolved dusB allele from population Ara+5 involves an A-to-T amino acid replacement at residue 315. All five evolved dusB alleles have mutations within a 45-bp region encoding the C-terminal region of DusB, and none of them appear to impact the dihydrouridine synthase activity (Savage et al. 2006).

In order to analyze the phenotypic effects of these alleles, we examined a set of isogenic strains made by allelic replacements in the ancestral background of strain REL606: 606fis−1, 606fis+1, 606fis−3, and 606fis−5, bearing the evolved fis alleles from Ara−1, Ara+1, Ara−3, and an in-frame fis deletion, respectively, and 606dusB+5 and 606dusB−5, which have the Ara+5 dusB evolved allele and an in-frame deletion of dusB, respectively. We also analyzed 5KdusBanc, in which the ancestral dusB allele replaced its evolved counterpart in a 5,000-generation clone sampled from the Ara+5 population.

The fis−1, fis+1, and dusB+5 Evolved Alleles Affect the Level of Fis Protein

We measured the effects of the various fis and dusB alleles on the level of Fis protein. Soluble proteins were extracted from early exponential-phase cultures, when the level of Fis is maximal, and immunodetection was performed using an antibody raised against Fis (fig. 2A). An anti-RpoA (2 subunit of RNA polymerase) antibody served as a control. The isogenic mutant strains (606fis−1, 606fis+1, 606fis−3, 606fis−5, 606dusB+5, and 606dusB−5), along with clones from all 12 populations at 20,000 generations, were compared with the ancestral strain REL606. Ancestral and evolved backgrounds gave qualitatively similar results for each fis and dusB allele when isogenic constructs were available (data not shown). Consistent with previous results (Crozat et al. 2005), the Ara−1 allele caused a roughly 3-fold reduction in the level of Fis, and a similar decline was observed for the Ara+1 allele (fig. 2A). In contrast, there were no comparable reductions for the Ara−3 and Ara−4 alleles, consistent with their mutations being located in the Fis coding sequence. Introduction of the evolved Ara+5 dusB allele into the ancestor also reduced the amount of Fis to about the same level as seen in the 20,000-generation Ara+5 clone. Moreover, deletion of dusB led to the absence of any detectable Fis protein, suggesting a positive regulatory role of DusB on fis expression. Complementation of the 606dusB strain with a plasmid overexpressing dusB restored the expression of fis (fig. 2A). This restoration, together with the fact that the dusB deletion was in-frame, eliminates potential polar effects that might have resulted from the dusB-fis operon structure.

A comparison of fis transcription levels in fis::lacZ chromosomal transcriptional fusions constructed in strains REL606 and 606dusB+5 revealed no obvious differences, indicating that the regulation of fis by DusB might occur at a posttranscriptional level (data not shown). However, deleting either fis (606fis−) or dusB (606dusB−) resulted in an approximately 5-fold increase in transcription of fis. This increase was probably a direct consequence of the absence of Fis protein in both strains, relieving the autorepression of Fis on its own transcription and also confirming the regulation of fis by DusB. The absence of an effect on transcription in the 606dusB+5 strain suggests that enough Fis
protein is still present to bind to the high-affinity sites in the fis promoter region.

In the light of these regulatory effects, the Fis levels measured in the 12 evolved clones agree well with the types of alleles they possess (fig. 2A). Five of them have mutations in dusB (Ara−6, Ara+3, Ara+4, Ara+5, and Ara+6), and all show reductions in Fis levels, as do two or three other clones with mutations in regulatory elements of the fis gene (Ara−1, Ara+1, and perhaps Ara+2). Two evolved clones with mutations in the fis coding sequence (Ara−3 and Ara−4), as well as two with no mutations in either fis or dusB (Ara−2 and Ara−5), express levels of Fis that are similar to the ancestral strain REL606.

Because the fis−1 allele affects the promoter region (fig. 1), we also examined its effect on transcription. Single-copy chromosomal lacZ transcriptional fusions were created with either the ancestral or the Ara+1 evolved fis promoter using bacteriophage λ derivatives (Simons et al. 1987). Each fusion was introduced into the λ att site of the ancestral REL606 chromosome. The β-galactosidase activities were then measured in exponential phase, during the transition into stationary phase, and in stationary phase (fig. 2B). The ancestral and evolved promoters exhibit similar dependence of fis transcription on growth rate, with maximum transcription during exponential growth followed by declining activity as cells transition into stationary phase. However, the activity of the evolved promoter was substantially reduced in all growth phases, consistent with the protein data (fig. 2A and B).

The fis−3 and fis−4 Evolved Alleles Affect the Activity of Fis Protein

We examined the effects of the Ara−3 and Ara−4 fis alleles by measuring in vivo the DNA-binding activities of Fis proteins using two compatible plasmids (Gosink et al. 1996). The expression plasmid contains either the fis gene or a variant under the control of the tac promoter, whereas the reporter plasmid carries the cat gene whose expression is repressed by Fis binding (Gosink et al. 1996). The cat gene encodes chloramphenicol acetyltransferase, which confers resistance to chloramphenicol, and its expression depends on modified lac−35 and −10 promoter elements. In particular, the lac promoter has been modified by introducing the UP elements of the rrnB1 promoter immediately upstream of the −35 box to strongly activate transcription and a strong Fis-binding site immediately downstream of the −10 box. Binding by Fis protein will repress cat transcription and thereby reduce the level of resistance to chloramphenicol of a strain carrying both plasmids. Moreover, the decrease in resistance should be proportional to the strength of Fis DNA binding.

The ancestral, Ara−3, and Ara−4 fis genes were each cloned into the expression plasmid under the control of P(tac). Each expression plasmid was introduced with the reporter plasmid into the ancestral strain carrying the in-frame fis deletion. All three strains were then plated onto two types of LB plates, one containing 100 μg/ml of chloramphenicol and the other containing 600 μg/ml of chloramphenicol; both types of plates also contained ampicillin to ensure retention of the expression plasmid. Relative survival was calculated as the ratio of colony number on plates with high and low chloramphenicol concentrations and used as an indirect measure of Fis binding, as previously described (Gosink et al. 1996). Survival was almost 9-fold higher when the expression plasmid carried the evolved fis allele from Ara−3 or Ara−4 (3.5% and 3.4%, respectively) compared with the ancestral gene (0.4%), indicating reduced DNA binding by the evolved Fis proteins.

To test this inference further, we purified the ancestral and evolved Fis proteins (Pan et al. 1996) and assessed their DNA-binding abilities by performing gel mobility shift assays (fig. 3). Three representative promoter sequences, all known to contain Fis-binding sites, were used as binding targets: the fis promoter (Mallik et al. 2004), the rrnB1 promoter (Ross et al. 1990), and the proP promoters (McLeod et al. 2002). For the ancestral protein, we observed a shift in electrophoretic mobility even at the lowest Fis concentration (0.2 nM), whereas retardation was maximal at the highest concentration used (fig. 3A). The multiple retarded bands indicate the presence of several Fis-binding sites (Xu and Johnson 1995). In contrast, both evolved Fis proteins exhibited reduced DNA-binding abilities, with higher concentrations (2–5 nM) necessary to produce a discernible shift (fig. 3A). Much higher concentrations (at least 20–30 nM) of the evolved proteins were also required for maximal retardation.

Binding specificity was checked by using 1) a 5- to 1,000-fold excess of an unlabeled 30-bp oligonucleotide containing a consensus Fis-binding site, 2) the same DNA fragment but with the Fis site mutated, and 3) an antibody against Fis in the DNA-binding reactions. The first assay resulted in a reversal of the binding observed in figure 3A; the second assay had no detectable effect on binding, and the third assay resulted in a supershift (data not shown). To compare more precisely the behavior of the ancestral and evolved Fis proteins, we also quantified the amount of DNA bound by Fis based on the gel mobility shift assays. Figure 3B shows a typical quantitation curve, whereas figure 3C provides relative dissociation constants calculated for each of the three Fis proteins and each of the three promoter regions. These values are 12–15 times higher for both evolved proteins compared with the ancestral protein, confirming their decreased DNA-binding ability.

Fitness Effects of the Evolved Alleles

We performed competition assays to estimate the fitness effects of two evolved fis alleles and one evolved dusB allele under the same conditions as the evolution experiment. All evolved alleles were introduced into the chromosome of the Ara− ancestral strain REL606, and these strains competed against the Ara+ ancestral strain REL607; the Ara marker is itself selectively neutral under these conditions (Lenski et al. 1991; Lenski 2004). Surprisingly, each evolved allele was either deleterious or neutral in the ancestral background (606dusB+; relative fitness 0.9482, t0 = 2.9232, 5 df, P = 0.0329; 606fis−3: 0.9535, t0 = 5.5278, 5 df, P = ...
dusB directly. Instead, 5K and 5K

6.9591, 5 df, \( P = 0.3546 \), based on a two-tailed \( t \)-test. Thus, the evolved \( \text{dusB} \) allele appears to be neutral based on a phylogenetic analysis of the distribution of the \( \text{DusB} \) and Fis proteins (Morett and Bork 1998). Sequence alignments of diverse \( \gamma \)-proteobacterial Fis proteins and \( \alpha \) - and \( \gamma \)-proteobacterial NtrC proteins indicate that Fis is probably derived from some ancestral version of the C-terminal DNA-binding domain of the \( \alpha \)-proteobacterial NtrC. Moreover, DusB has significant similarity to NifR3, encoded by a gene located in the same operon as ntrC

0.0027; and 606\(^{\text{dusB-5}+} \) 1.0002, \( t_s = 0.1205, 11 \text{ df, } P = 0.9063 \); in each case, \( H_0 = 1 \) was tested using a two-tailed \( t \)-test. These outcomes contrast with the small, but significant, benefit measured for the evolved Ara

Phylogenetic Confirmation of the Functional Link between \( \text{dusB} \) and Fis

An earlier study suggested a functional link between \( \text{dusB} \) and Fis based on a phylogenetic analysis of the distribution of the \( \text{DusB} \) and Fis proteins (Morett and Bork 1998). Sequence alignments of diverse \( \gamma \)-proteobacterial Fis proteins and \( \alpha \) - and \( \gamma \)-proteobacterial NtrC proteins indicate that Fis is probably derived from some ancestral version of the C-terminal DNA-binding domain of the \( \alpha \)-proteobacterial NtrC. Moreover, DusB has significant similarity to NifR3, encoded by a gene located in the same operon as ntrC

To summarize our results to this point, we observed strong genetic parallelism during the evolution experiment, with Fis and \( \text{dusB} \) repeatedly being targets of natural selection. This parallelism was achieved, however, through different mutations that affected all levels of regulation: transcription, translation, protein activity, and regulation via a newly discovered regulator of \( \text{fis} \), the \( \text{DusB} \) protein. Despite these mutational differences, however, all of the evolved variants are phenotypically similar insofar as they cause a reduction in either the amount or the activity of Fis protein. However, we were unable to measure significant fitness advantages for the mutations we tested, perhaps because benefits are very small or depend subtly on context, although the genetic and phenotypic parallelisms strongly imply that they were beneficial during the long-term evolution experiment.
Evolutionary Changes in DNA Superhelicity  

in several α-proteobacteria. These earlier sequence analyses further suggested that the dusB-fis operon resulted from some rearrangement involving horizontal transfer into a lineage ancestral to the γ-proteobacteria (Morett and Bork 1998). These patterns were offered as evidence for some interaction between the proteins to “carry out an as yet unknown function.”

This previous study on the association of Fis and DusB was performed in 1998 using fewer than 30 proteins in the alignments. We revisited the phylogenetic association by analyzing the fis and dusB sequences in 323 bacteria with available genome sequences. The fis gene was not found at all outside the proteobacteria. The association of dusB and fis in the same operon was widely observed in both the γ-proteobacteria (53 of 79 genomes examined) and the β-proteobacteria (22 of 23 genomes). By contrast, this same association was rare in the α-proteobacteria (1/46 genomes) and δ-proteobacteria (1/11 genomes). These observations reinforce the functional association between fis and dusB. They also suggest that the dusB-fis operon arose before the γ- and β-proteobacteria diverged, whereas the earlier phylogenetic analysis suggested a more recent origin within the γ-proteobacteria (Morett and Bork 1998).

Test of the Effect of Genetic Context on the Evolution of Increased Supercoiling

Unlike the other ten populations, two lines, Ara+1 and Ara+3, had not evolved increased supercoiling within 20,000 generations (Crozat et al. 2005), nor had either line substituted any mutations in the topA gene (table 2), and yet they had improved fitness similar to the other populations (Cooper and Lenski 2000). We asked, therefore, whether mutations elsewhere in the genome had eliminated the advantage of increased DNA supercoiling. We moved the topA allele that had been substituted in population Ara−1 (Crozat et al. 2005), and which strongly increases both fitness (+13%) and DNA superhelicity (from −0.066 to −0.075), into the genome of an evolved clone isolated at generation 2,000 from population Ara+1. This clone, called 1158C, still carries the ancestral topA allele, and it shows only a slight increase in DNA supercoiling (−0.069) relative to the ancestor (Crozat et al. 2005). We then compared the supercoiling and fitness levels of 1158C and its 1158CtopA1 counterpart. Replacing the ancestral topA with the evolved topA1 allele increased supercoiling in the 1158C clone to the same level as generated by the topA1 allele in the ancestral genetic background (fig. 4A). The evolved topA1 allele also significantly increased fitness in both the ancestral (Crozat et al. 2005) and the evolved Ara+1 backgrounds (fig. 4B).

Thus, in at least one lineage that evolved little or no increase in DNA supercoiling, the introduction of a beneficial supercoiling-related mutation from another line generated similar phenotypic effects, at least at the 2,000-generation time point that we tested. Therefore, the absence of an altered topA in that particular population does not appear to be the result of a very early substitution of some other mutation that eliminated the beneficial effect of topA.

Discussion

In a previous study, parallel phenotypic increases in DNA supercoiling were observed in 10 of 12 evolving E. coli populations (Crozat et al. 2005). In one population that evolved two successive increases in DNA supercoiling, two mutations were discovered, one in topA and the other in fis, and together, they were shown to be responsible for the observed changes in DNA topology in that population.
Both mutations were also demonstrated to increase fitness, indicating that supercoiling was a target of selection in the experimental environment (Crozat et al. 2005). More recently, whole-genome sequencing of several clones isolated from this same population through 20,000 generations revealed no other mutations in genes known to control DNA topology (Barrick et al. 2009).

Here, we extend the investigation of the DNA topology-associated regulatory network to all 12 populations. We discovered substantial parallelism at both the genetic and the molecular levels. Sequencing nine DNA topology-related genes encoding topoisomerases and histone-like proteins suggested that only two of them, topA and fis, had evolved in many or most of the populations. We also sequenced eight other genes in the same operons as the nine known topology-related genes and discovered that one of them, dusB, had also changed in several populations and, moreover, was involved in regulating DNA topology. All of the mutations in these three genes were fixed in their respective populations. In contrast to this parallelism at the level of the affected genes, however, there was some divergence in the mode of action of the substituted alleles. In particular, phenotypic analyses revealed that some of the mutations—those in the promoter of the dusB-fis operon, in the fis RBS, and in the coding sequence of dusB—reduced the level of Fis protein, whereas those in the coding region of fis decreased the protein’s activity.

The three genes identified here as the targets of selection on DNA topology—topA, fis, and dusB—represented only 20% of our sequencing effort, yet they harbored 16 of the 27 mutations found in the 17 genes sequenced. Moreover, all but 1 of the other 11 mutations occurred in the four populations that had evolved mutator phenotypes, which are known to accumulate neutral and mildly deleterious mutations at much higher rates (Snigowski et al. 1997; Cooper and Lenski 2000; Funchain et al. 2000; Giraud et al. 2001; Lenski et al. 2003; Barrick et al. 2009). In previous work on the genetic basis of adaptation in these same lines, but using different genes, mutations under positive selection were shown to be overdispersed (Woods et al. 2006); for example, each line substituted one nonsynonymous mutation in a particular gene, but none harbored two or more substitutions in the same gene. This nonrandom distribution implies that one mutation suffices to provide a benefit, whereas a second one in the same gene yields little or no further advantage. With respect to fis, we observed that 5 populations had single substitutions in this gene, which is consistent with the overdispersed distribution, although the pattern is not significant given only five mutations in 12 populations. However, we also noticed that five lines each had single mutations in dusB, which occurs in the same operon as fis. Moreover, no line had mutations in both fis and dusB, such that the ten mutations total in this operon were maximally, and significantly, overdispersed among the 12 populations. This pattern led us to hypothesize that these two genes interact in some way, such that a population could benefit similarly from a mutation in either fis or dusB. Our genetic experiments and molecular assays demonstrated that the DusB protein regulates expression of fis and confirmed that some mutations in dusB had similar functional effects to certain mutations in fis and its known regulatory elements. These experiments showed for the first time a regulatory function of DusB on the fis gene in addition to its previously described dihydrouridine synthase activity (Bishop et al. 2002). Our phylogenetic analysis of the linkage between dusB and fis also implies that DusB has a similar regulatory function in many other species besides E. coli.

Despite propagating the same ancestral strain in identical environments, all 12 lines have followed different evolutionary trajectories. In some cases, certain of the early steps have closed off particular avenues or opened new pathways for adaptation (Cooper et al. 2003; Blount et al. 2008); in other cases, the same opportunity may exist until each population has solved the challenge in a similar way. With respect to DNA topology, two lines had not yet changed even after 20,000 generations. When we moved a topA allele from a line that had evolved in this respect to a 2,000-generation clone from one of the populations that had not changed, that allele was also beneficial in this other genetic background, indicating that the potential for adaptive evolution of this trait had persisted at least to that point in time. More generally, despite parallelism at the level of genes, there was substantial allelic diversity, with no two evolved alleles identical at the nucleotide level. In particular, different mutations in fis and dusB affected different aspects of fis expression—transcription, translation, and protein activity—though they had similar effects because they all reduced either the amount or the activity of Fis. The mutation in population Ara-1 affects the promoter (Walker et al. 1999) and thereby reduces transcription. The mutation in Ara-1 affects the RBS element, reducing complementarity between the translational initiation region of fis messenger RNA (mRNA) and the 3′ end of 16S rRNA (Owens et al. 2004). The mutation in Ara-4 is a 1-bp insertion, leading to a change in the C-terminus of the protein such that each monomer is 103, rather than 98, residues long. This mutation modifies the D helix of Fis, which affects the DNA-binding helix-turn-helix (HTH) motif (Yuan et al. 1991); important residues are also absent from the evolved protein, which might explain its reduced DNA-binding activity. The evolved Ara-3 allele, by contrast, replaces the Y51 residue with an S residue, and the Y51 residue constitutes an important part of a hydrogen-bonded network that is involved in both the HTH DNA-binding motif and the Fis dimerization (Yuan et al. 1991). The dusB mutation in population Ara+5 disrupts positive regulation of Fis by DusB at the translational level. DusB is known to modify tRNAs (Bishop et al. 2002), and so we hypothesize that DusB might exert its regulatory effect by binding the dusB-fis mRNA, thereby facilitating the translation of fis. However, additional experiments are needed to test this hypothesis and further investigate the functional relationship between DusB and Fis. More generally, these evolved fis and dusB alleles, by reducing either the amount or the activity of Fis, may lead to an
increase in the level of DNA superhelicity by partially relieving the repression Fis has been shown to exert on the gyrAB genes that encode gyrase (Schneider et al. 1999).

Two interesting comparisons can be made with other evolution experiments that have been performed with various organisms and under diverse conditions. First, two of the main bacterial global regulatory networks, DNA superhelicity (Crozat et al. 2005, this work) and the stringent response (Cooper et al. 2003; Pelosi et al. 2006), have changed in many of the replicate populations in this long-term experiment with E. coli, as have other genes that may have profound metabolic effects (Philippe et al. 2007). More generally, changes in regulatory and metabolic networks appear to be a common and an important feature of laboratory evolution experiments with microbes, including E. coli propagated in other environments (Herring et al. 2006; Ferenci 2008; Conrad et al. 2009), as well as many other bacteria (Velicer et al. 2006; Bantinaki et al. 2007; Gresham et al. 2008) and yeast (Rosenzweig et al. 1994; Ferea et al. 1999; Gresham et al. 2008; Lang et al. 2009). These results indicate the important role of regulatory networks in sustaining the evolvability of living organisms.

Second, parallel evolution across replicate populations has been reported in many of these experiments, although the functional level and quantitative extent of parallelism vary between different systems. At one extreme, replicate populations of bacteriophage ΦX174 exhibited striking parallelism even at the DNA sequence level, with about half of all the fixed mutations identical across two replicate populations (Wichman et al. 1999). By contrast, in several studies of evolving bacteria, including this one, replicate populations frequently show parallelism at the level of genes and cellular functions but rarely at the level of the mutations themselves (Herring et al. 2006; Woods et al. 2006; Bantinaki et al. 2007; Barrick et al. 2009; Conrad et al. 2009). Woods et al. (2006) have hypothesized that the smaller size and lower complexity of viral genomes relative to bacteria might explain their greater propensity for parallel sequence evolution.

The parallel evolution that we have observed at the level of DNA superhelicity and the underlying genes provides strong evidence that the corresponding mutations had beneficial fitness effects, especially in the absence of any specific sequence motifs that might suggest hypermutability. Indeed, this interpretation has been confirmed for several other genes in these same evolving population system by constructing and competing isogenic strains that differ only by the alleles in question (Cooper et al. 2001, 2003; Pelosi et al. 2006), including previous studies of topA and fis alleles that changed supercoiling in one of the populations (Crozat et al. 2005). Yet, despite the signature of positive selection in the genetic and functional parallelism, two other fis and one dusB evolved alleles that we tested here were neutral or even slightly deleterious when moved into the ancestral background.

We do not know why the evolved alleles tested in this study failed to show the expected benefits, but we can suggest three possible explanations. One possibility is a methodological problem during the construction of the isogenic strains, in particular the possible introduction of some secondary mutation that arose during the genetic manipulations. Most mutations are deleterious, and so any such random mutation would be far more likely to reduce fitness than to increase it (hence, such problems are unlikely to explain beneficial effects when those are observed). A second possibility involves subtle differences in the environmental conditions used to assess fitness effects and those that existed during the evolution experiment itself. In particular, fitness assays involve pairwise competitions in which the strains have equal initial frequency. By contrast, many genotypes were often present during the evolution experiment (Papadopoulos et al. 1999; Barrick and Lenski 2009), and some of them have subtle fitness advantages only when they are in the minority (Elena and Lenski 1997; Rozen et al. 2005).

A third possibility is that evolved alleles may be beneficial in some genetic contexts but not in others. In particular, an evolved allele might be beneficial only in association with one or more mutations that arose previously, in which case the allele in question would not be advantageous when tested in the ancestral strain. This possibility may be particularly relevant here because pervasive epistatic interactions involving global regulatory networks have been demonstrated in the long-term populations (Cooper et al. 2008). Moreover, it is known that the stringent response and DNA superhelicity regulatory networks are intertwined (Travers and Muskhelishvili 2005), and both have been evolving in these populations (Cooper et al. 2003, 2008; Crozat et al. 2005; Philippe et al. 2007). Also, different evolved alleles may have different pleiotropic effects, so that compensatory changes might be specific to a given allele, thereby giving rise to epistatic interactions with mutations in some other genes. Given these potential subtleties and complications, it would be interesting to reconstruct and analyze in detail the effects of all of the mutational steps in one or more of these populations, though that remains a daunting challenge in light of the many mutations and the difficulty of ensuring strict isogenic comparisons. For now, we sought to test this epistatic hypothesis by replacing the evolved dusB allele (which had appeared to be neutral in the ancestral background) in an evolved clone with the ancestral dusB allele. The result was inconclusive at best, however, with the evolved allele again having a negative, albeit not significant, effect in this context.

In conclusion, a central goal of evolutionary biology is to characterize the often complex relationships between genotype, phenotype, and fitness. We previously showed that mutations in topA and fis were responsible for changes in DNA superhelicity in 1 of 12 populations in a long-term experiment with E. coli (Crozat et al. 2005). Here, we extended these results to the other populations in that experiment by showing that mutations in topA, fis, and a third gene, dusB, caused parallel changes in superhelicity in most of them. The dusB gene was not previously known to have any role in controlling DNA topology prior to our analyses. We showed that DusB exerts its effect on
superhelical by regulating the expression of the Fis protein. However, it remains unknown how a reduction in the amount or activity of Fis leads to the observed changes in supercoiling and if other yet unknown genetic changes may also be involved. Ongoing genome sequencing and global expression analyses of isolates from the long-term evolution experiment should provide new insights into these questions. It will be especially useful to obtain global transcription profiles of strains that are isogenic except for the various fis and dusB mutations to evaluate the impact of those mutations on other genes that may affect such phenotypes as DNA superhelicity and fitness. Comparisons of these profiles across the ancestral and evolved backgrounds should also give insights into other mutations that may interact with the topology-related alleles. Histone-like proteins such as Fis are involved in both global transcription control and global chromosomal organization, and thus, it would also be informative to obtain and compare the complete maps of Fis-binding sites in the ancestral and evolved genomes. These maps could be obtained by chromatin immunoprecipitation coupled with microarray technology or ChIP-chip (Grainger et al. 2006).Taken together, such analyses would provide valuable insights into the role of changing gene regulatory networks in the relationship between genotype, phenotype, and fitness.

Acknowledgments

The authors thank Mohamed-Ali Hakimi (Laboratoire Adaptation et Pathogénie des Micro-Organismes, France) for help in Fis purification, Akira Ishihama (Nippon Institute for Biological Sciences, Japan) and Michael Cashel (National Institutes of Health, USA) for the anti-Fis and anti-RpoA antibodies, respectively, and Tamas Gaal and Richard Gourse (University of Wisconsin, USA) for plasmids pKG13 and pKK223-3. This work was supported by grants from the French Centre National de la Recherche Scientifique (to D.S., BDI fellowship to E.C.), Université Joseph Fourier (to D.S.), Agence Nationale de la Recherche Program “Génomique Microbienne à Grande Echelle” (ANR-08-GENM-023 to D.S.), the United States National Science Foundation (DEB-9981397 and DEB-0515729 to R.E.L.), and the United States Defense Advanced Research Projects Agency “Fun Bio” Program (HR0011-09-1-0055 to R.E.L.).

References


