ISSN 0026-8933, Molecular Biology, 2007, Vol. 41, No. 5, pp. 827–835. © Pleiades Publishing, Inc., 2007.
Original Russian Text © L.V. Sycheva, E.A. Permina, M.S. Gelfand, 2007, published in Molekulyarnaya Biologiya, 2007, Vol. 41, No. 5, pp. 908–917.

GENOMICS. TRANSCRIPTOMICS. PROTEOMICS

UDC 577.214.4

Taxon-Specific Regulation of the SOS Response in γ-Proteobacteria

L. V. Sycheva[a], E. A. Permina[b], and M. S. Gelfand[a, c]

[a] *Department of Bioengineering and Bioinformatics, Moscow State University, Moscow, 119992 Russia; e-mail: lada.sychova@gmail.com*

[b] *State Research Center GosNIIgenetika, Moscow, 113545 Russia*

[c] *Kharkevich Institute of Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994 Russia*

Received November 7, 2006

Accepted for publication January 22, 2007

Abstract—The SOS response is a cascade of consecutive reactions induced by cell DNA damage. The genes directly involved in these reactions are regulated by LexA, which binds to specific nucleotide sequences in their upstream regions. The presence of such a sequence in the regulatory gene region can be used as a criterion to identify the genes potentially involved in the SOS response. A study was made of the genes whose regulation is specific to particular taxa (Enterobacteriales, Pasteurellales, Vibrionales, Pseudomonadales, and Alteromonadales). Some of the genes identified have not been implicated in the SOS response as yet but have a conserved LexA-binding site in the regulatory region and perform a function probably associated with the cell response to DNA damage. These genes include *mfd*, whose product facilitates DNA repair when transcription is arrested because of DNA damage; *VC0082*, coding for recombinase; and *VP2449*, which is responsible for xenobiotic resistance. The composition and evolution of the LexA regulon in γ-proteobacteria are considered.

DOI: 10.1134/S0026893307050160

Key words: SOS response, LexA, γ-proteobacteria, transcription regulation, comparative genomics

INTRODUCTION

The SOS response is a cascade of consecutive reactions induced by DNA damage [1]. Defais et al. [2] were the first to suggest such a system, and Radman et al. [3] verified and developed their hypothesis. The SOS response protects the cell from DNA damage by a severe chemical exposure, irradiation, high pressure [4], or temperature. Several cell mechanisms are known to maintain the genome stability: recombination repair, mutagenic repair, excision repair, mismatch repair, and DNA photolyase repair [5]. The SOS response has been well studied in *Escherichia coli*, including both the regulatory mechanisms and the genes whose products are involved in repairing DNA lesions [6].

Normally, LexA represses all SOS-response genes by binding to a specific sequence, known as an SOS box (consensus TACTGTATATATATACAGTA [7]), in their 5'-noncoding regions. Genes of the SOS regulon vary in the distribution of LexA-binding sites and their affinity for LexA [8]. When DNA is damaged, RecA binds to its single-stranded regions and forms nucleoprotein filaments, being converted into an active form, RecA*. As a result of the RecA*–LexA interaction, LexA is proteolytically cleaved at Ala84–Gly85 and the SOS-response genes are derepressed.

When the consequences of damage are eliminated, the RecA* pool is reduced and the noncleaved LexA pool is restored. The RecA pool is decreased via LexA-dependent repression of *rec*A, allowing feedback.

The LexA regulon includes the genes responsible for break repair in the daughter and double DNA strands and recombination (*rec*A); the genes for polymerases (*umu*DC and *din*P), recombinases (*rec*A and *rec*N), nucleases, helicase (*uvr*D), and a cell division inhibitor (*sul*A); and the genes involved in excision repair (*uvr*AB) [6]. The single-stranded DNA-binding protein (Ssb) is among the first to be involved in the SOS response, binding to single-stranded DNA regions [9]. This binding is necessary for subsequent replication, since Ssb is a component of the DNA polymerase complex. Other components of the complex (polymerase V) are encoded by *umu*D and *umu*C. The polymerase V genes are changed by nonorthologous substitutions in some cases. For instance, the *Caulobacter crescentus* operon consisting of *imu*A, *imu*B, and *dna*E2 codes for the catalytic subunit of polymerase III and is regulated by RecA [10].

Table 1. Methods predicting the binding sites for gene expression regulators

Total genes	Searching tool	Nucleotide sequence	Reference
69	Consensus sequence	TACTG-$(TA)_5$-CAGTA	[19]
39	Consensus sequence	CTG-N_{10}-CAG	[13]
19	Consensus sequence, position weight matrix	TACTGT-$(AT)_2$-ACAT-A/C-CAG-T/C-A	[17]
54	Position weight matrix	TACTG-$(TA)_5$-CAGTA	This work

The complex enzyme formed by the products of *umu*D and *umu*C, RecA, and Ssb functions more rapidly than vegetative polymerase but makes more mistakes. In *Pseudomonas putida*, mutagenesis associated with polymerase IV is not related to SOS-response induction because transcription from the promoter of *din*B, coding for polymerase IV, is independent of the SOS activity of the cell. A region with a mistake in the nucleotide sequence is excised during the SOS response by nuclease, a complex enzyme consisting of UvrA and UvrB. RecA, RecN, and helicase UvrD are responsible for DNA recombination repair.

DNA damage can cause cell death. The critical factor is whether the SOS system is capable of restoring the genetic material. The mechanism triggering the SOS response is active throughout the life cycle of the cell, in particular, during cell division, which can be arrested by SulA as a result of the SOS response until DNA lesions are eliminated.

LexA orthologs are found in various taxa. In *Bacillus subtilis*, the SOS system is regulated by DinR, an ortholog of *E. coli* LexA. A comparison of DinR with its homologs of proteobacteria and Gram-positive bacteria has revealed conserved regions in the C-terminal domain, which is presumably responsible for autocatalytic cleavage [1]. A helix–turn–helix (HTH) domain is also conserved in the regulatory proteins of γ-proteobacteria and Gram-positive bacteria but is insufficient for the preservation of the signal motif. Like *E. coli* LexA, DinR is capable of autocatalytic cleavage and represses *rec*A by binding to its promoter region [11]. The *B. subtilis* DinR-binding site (Cheo box) is palindromic; its consensus (CGAACATATGTTC) differs from that of the proteobacterial SOS box [12].

Analysis of the LexA regulon in α-proteobacteria by experimental and in silico methods has implicated several new genes in the SOS response [13]. In particular, these are *par*E, coding for subunit B of DNA topoisomerase IV, and *com*M, whose product acts as a Mg^{2+} chelatase and regulates the polymerase function during the SOS response. The conserved tandem repeat GTTC-N_2-GTTC of the LexA-binding site has been found in the *Rhodobacter, Sinorhizobium, Agrobacterium, Caulobacter*, and *Brucella* genomes [14].

The possibility to predict binding sites has been studied with several methods such as the consensus construction algorithm [15], mathematical expectation maximization algorithm [16], oligonucleotide frequency analysis [17], and the Gibbs sampling algorithm [18] (Table 1).

A consensus sequence can be used to search for regulatory motifs [19]. A nucleotide sequence in question is compared with the consensus by means of the heterology index (HI). Sequences with HI < 15 have a high probability of playing a regulatory role and binding LexA.

A three-step algorithm has been developed to search for regulatory motifs [20]. Genomic DNA is searched for sequences matching the regulatory motif. The findings are passed through a recursive filter and a consensus matrix is constructed. The motifs with HI < 8 are selected.

At the last step, the program automatically extracts the functional annotations of the potential regulon members from GenBank, using the TBLASTN server.

An algorithm has been developed to identify the interconnected metabolic modules, which represent regulons at the genome level [21]. The regulon genes are predicted and bacterial genomes analyzed using the information theory methods, programming, and Bayesian statistics. The results obtained for new bacterial genomes are deposited in RegulonDB (http://www.cifn.unam.mx/Computational_Genomics/regulondb/). Regulatory signals are sought with a consensus matrix, which is used to construct a weight matrix by the Patser program.

We have previously studied the regulon core in several γ-proteobacteria (unpublished data). The regulon is probably conserved among phylogenetically close organisms. Based on this assumption, the regulon can be studied in the genomes of organisms whose potential regulon members have not been experimentally tested for regulation as of yet.

The objective of this work was to examine the genes whose regulation is specific to the genomes of a particular taxonomic group; that is, we were aimed at studying the total LexA regulon in γ-proteobacteria. Our findings can be used to analyze the evolution of the LexA regulon.

EXPERIMENTAL

To analyze the γ-proteobacterial SOS response, we compared the regulation of genes belonging to the LexA regulon [22]. First, we verified the regulatory sites in the upstream regions of the genes whose LexA-dependent regulation was experimentally verified. Then, the total set of LexA-regulated genes was used to select the genes that have the regulatory sites only in the genomes of a particular phylogenetic group (i.e., the genes with taxon-specific regulation). Apart from the presence of a potential regulatory site in the upstream region, a gene was assigned to the set with taxon-specific regulation on evidence of the ratio (50%) between the minimal number of genomes where the gene does have the regulatory site and the total number of completely sequenced genomes in the group, as well as the gene function.

Potential regulatory binding sites were compared using the GenomeExplorer software program [23].

Searching nucleotide sequences for a signal, it is expedient to consider the regions where the signal is more likely to occur. The regulatory function is usually performed by sequences of the upstream region. Hence, we searched the regions from –200 to 50 relative to the annotated gene start.

The weight of a site was computed using a position weight matrix (PWM) (Table 2), which was constructed with the SignalX program of the GenomeExplorer software package by the following equation:

$$W(b, k) = \log[N(b, k) + 0.5]$$

$$- 0.25 \sum_{\mathrm{i = A, C, G, T}} \log[N(i, k) + 0.5],$$

where $N(b, k)$ is the occurrence of nucleotide b in position k.

The weight of a putative site was computed as a sum of the weights of the nucleotides in the corresponding positions:

$$Z(b_1 \ldots b_k) = \sum_{\mathrm{k = 1 \ldots k}} W(b_k, k),$$

where k is the site length. The base of the logarithm was selected so that the Z distribution had a mathematical expectation of 0 and a variance of 1 in a set of random sequences. Weight Z characterizes the significance of an individual site.

Sites with a weight of no less than 3.75 were considered. With these threshold and search parameters, 92 potential sites were found in the *E. coli* genome.

Multiple sequence alignments and phylogenetic trees were constructed using the ClustalX [24] and Phylip [25] software programs. The GeneMaster software program was used to plot the phylogenetic trees.

Table 2. Position weight matrix employed in searching for LexA-binding sites

A	C	G	T
−0.01	0.03	−0.15	0.12
0.23	−0.05	−0.01	−0.17
−0.15	0.46	−0.15	−0.15
−0.15	−0.15	−0.15	0.46
−0.15	−0.15	0.46	−0.15
−0.13	−0.13	−0.05	0.30
0.28	−0.07	−0.14	−0.07
−0.22	0.07	−0.22	0.37
0.23	−0.31	0.10	−0.03
0.08	−0.07	−0.19	0.18
0.15	−0.14	−0.05	0.04
0.01	−0.01	−0.11	0.11
0.23	−0.05	−0.17	−0.01
0.05	0.03	−0.21	0.13
0.22	0.15	−0.31	−0.07
−0.15	0.46	−0.15	−0.15
0.46	−0.15	−0.15	−0.15
−0.15	−0.15	0.46	−0.15
−0.11	0.01	−0.11	0.21
0.12	−0.15	0.01	0.01

We examined 18 complete genomes of γ-proteobacteria from five subgroups: Enterobacteriales (*E. coli* (ECC) [25, 26], *Salmonella typhi* (STY) [28], *S. typhimurium* (STM) [29], *Shigella flexneri* (SFX) [30], *Yersinia enterocolitica* (YEN) [31], *Y. pestis* (YPE) [26], and *Photorhabdus luminescens* (PLU) [32]), Pasteurellales (*Haemophilus ducreyi* (HDU) [33], *H. influenzae* (HIN) [34], and *Pasteurella multocida* (PMU) [35]), Vibrionales (*Vibrio cholerae* (VCH) [26], *V. parahaemolyticus* (VPA) [36], *V. vulnificus* (VVU) [37], and *Photobacterium profundum* (PPR) [38]), Pseudomonadales (*Pseudomonas aeruginosa* (PAE) [26], *P. putida* (PPU) [38], and *P. syringae* (PST) [38]), and Alteromonadales (*Shewanella oneidensis* (SON) [39]).

To search for homologs in GenBank and TCDB [40], the BLASTP software program was used with the default parameters [41]. In addition, we used the Clusters of Orthologous Groups (COG) database [42].

RESULTS

Regulon Core

The core of a regulon is formed by the genes that are regulated in phylogenetically distant organisms and are certainly regulated in a group of phylogeneti-

Table 3. Regulon core and genes with potential taxon-specific regulation in the genomes of Enterobacteriales

Gene*	Annotation	ECC	STY	STM	SFX	YEN	YPE	PLU
*lex*A	SOS regulator	+	+	+	+	+	+	+
*rec*A	SOS coregulator	+	+	+	+	+	+	+
*rec*N	DNA repair	+	+	+	+	+	+	+
*ruv*A	Holliday junction DNA helicase	+	+	+	+	+	–	+
*uvr*A	Endonuclease ABC subunit A	+	+	+	+	–	–	–
*uvr*B	Endonuclease ABC subunit B	+	+	+	+	+	+	+
*uvr*D	DNA helicase II	+	+	+	+	+	+	+
*umu*D	Regulator of the polymerase complex	+	+	+	+	+	0	0
*din*D	DNA damage-inducible protein D	+	0	0	+	0	0	0
*din*G	ATP-dependent helicase	+	0	+	+	–	–	–
*din*I	DNA damage-inducible protein I	+	+	+	+	+	+	+
*fts*K	Cell division protein	+	+	+	+	+	+	+
*sul*A	Cell division inhibitor	+	+	+	+	+	+	+
ssb	Single-stranded DNA-binding protein	+	+	+	+	+	+	+
*din*P	DNA polymerase IV	+	+	+	+	+	+	+
*ycc*R	DNA-transforming protein	+	+	+	+	+	+	0
*ydj*Q	Nuclease subunit of the exonuclease complex	+	+	+	+	–	–	–
*ots*B	Osmotic pressure regulator	+	+	+	+	0	0	0
*sbm*C	Gyrase inhibitor	+	+	+	0	+	0	0
*yfi*K	Threonine transporter	+	+	+	+	0	0	0
*ygj*F	G/U mismatch-specific DNA glycosylase	+	+	+	0	0	0	0
*ydj*M	Membrane-associated metal-dependent hydrolase	+	+	+	+	+	+	–

Notes: Here and in Tables 4–6, the genomes are designated as in Experimental. (+), the gene has a potential regulatory site; (–), the potential regulatory site is lacking; 0, the ortholog is absent from the genome.

* The genes are designated as in *E. coli.*

cally related organisms. The regulator-binding site is highly conserved among such genes.

The most conserved core of the SOS regulon includes *lex*A, *rec*A, and *rec*N. These genes have a LexA-binding site in all genomes examined.

Enterobacteriales

In the *E. coli, S. typhi, S. typhimurium, S. flexneri, Y. enterocolitica, Y. pestis*, and *P. luminescens* genomes, LexA-dependent regulation was demonstrated for *lex*A, *rec*A, *rec*N, *uvr*B, *uvr*D, *din*I, *fts*K, *sul*A, *ssb*, and *din*P (Table 3), which have been experimentally assigned to the SOS regulon.

Several genes of the regulon core have an upstream regulatory site only in some representatives of the group. The regulatory site upstream of *ruv*A was found in six genomes but not in the *Y. pestis* genome. The upstream regions of *uvr*A and *din*G is lacking in the *Y. enterocolitica, Y. pestis*, and *P. luminescens* genomes. The *S. typhi* genome lacks *din*G.

The *Y. pestis* and *P. luminescens* genomes lack *umu*D. The regulatory site upstream of *din*D occurs in two genomes, while the gene itself is absent from the *S. typhi, S. typhimurium, Y. enterocolitica, Y. pestis*, and *P. luminescens* genomes.

The *ydj*M gene is a new potential member of the SOS regulon with taxon-specific regulation (Fig. 1). The LexA-binding site in its upstream region was found in all but one (*P. luminescens*) organism of the group. The gene belongs to the cluster of orthologous genes coding for membrane-associated metal-dependent hydrolase (COG1988).

In six genomes, a regulatory motif was found for *ycc*R, which codes for a DNA-transforming protein (Fig. 2). This gene belongs to a cluster of orthologous genes coding for regulators of competent cell-specific genes (COG3070). The gene is absent from the *P. luminescens* genome.

In four genomes, LexA presumably regulates *ots*B (osmotic pressure regulator), *ydj*Q (nuclease subunit

```
STM | ydjM       TCCGCCTACTGTATAAAAACCCTATACTGTATGAATTGACAGTT-----
STY | STY1790   TCCGCCTACTGTATAAAAACCCTATACTGTATGAATTGACAGTT-----
ECC | ydjM       TCCGTGCACTGTATAAAAACCCTATACTGTACGTATCGACAGTT-----
SFX | S1619     TCCGTGCACTGTATAAAAATCCTATACTGTACGTATCGACAGTT-----
YPE | YPO1717  GTTCACGCCAAGAAAAATCTCATATACTGGATAAATCAACAGCTACAAA
YEN | 001_1502 ATTCATGTCATTAAAAATCTCATATACTGGATAAATCAACAGCTACAGA
                     *   * ***   * ******* *   **  **** *
```

Fig. 1. Alignment of the *ydj*M upstream regions. One putative LexA-binding site is in bold and another one is underlined. Here and in Figs. 2–7, the positions occupied by the same nucleotide in all sequences of an alignment are indicated with asterisks.

```
ECC | yccR      TGTGAGTTACTGTATGAATGTACAGTACATCCAGTGACGAC
SFX | S1025     TGTGAGTTACTGTATGGATGTGCAGTACATCCAGTGACAAC
STM | yccR      TGTGAGTTACTGTATATTCATACAGTAC-CCCTGTGGCGAT
STY | STY1094   TGTGAGTTACTGTATATTCATACAGTAC-TCCTGTGGCGAT
YEN | 001_1314 TGTGGTGTCCTGTATATATATACAGTAGTCACTGTT-TGTT
YPE | YPO1437  TGTAGTGTGCTGTATGTATATACAGTAGTCACTGTT-TAAT
               ***    * ******     * *****    * **
```

Fig. 2. Alignment of the *ycc*R upstream regions. Here and in Figs. 3–7, a putative LexA-binding site is in bold.

```
STY | STY1804  ATACTGGATGAATAACCAGTTAA
STM | STM1309  ATACTGGATGAATAACCAGTTAA
ECC | ydjQ     ACACTGGATAGATAACCAGCATT
SFX | S1602    ACACTGGATAGATAACCAGCATT
               * *******  ********
```

Fig. 3. Alignment of the *ydj*Q upstream regions.

```
ECC | yfiK      TGTCGGGTACTGTCTACCAAAACAGAGGAGATA
SFX | yfiK      TGTCGGGTACTGTCTACCAAAACAGAGGAGATA
STY | STY2838  TATCGGGTACTGTCTGCTAAAACAGAGGAGATG
STM | yfiK      TATCGGGTACTGTCTGCTAAAACAGAGGAGATG
               * ************* * ***************
```

Fig. 4. Alignment of the *yfi*K upstream regions.

of the exonuclease complex, COG0322) (Fig. 3), *sbm*C (gyrase inhibitor, COG3449), and *yfi*K (transport protein, COG1280) (Fig. 4).

Three genomes (*E. coli, S. typhi*, and *S. typhimurium*) have a regulatory site upstream of *ygi*F, coding for G/U mismatch-specific DNA glycosylase.

Table 4. Regulon core and genes with potential taxon-specific regulation in the genomes of Pasteurellales

Gene*	Annotation	HIN	HDU	PMU
HI0749	*lex*A, SOS regulator	+	+	+
HI0600	*rec*A, SOS coregulator	+	+	+
HI0070	*rec*N, DNA repair	+	+	+
HI0313	*ruv*A, Holliday junction DNA helicase	+	+	+
HI0250	*ssb*, single-stranded DNA-binding protein	+	+	+
HI1188	*uvr*D, DNA helicase II	+	+	+
HI0249	*uvr*A, endonuclease ABC subunit A	+	–	+
HI1258	*mfd*, transcription–repair coupling protein	+	–	+

* The genes are designated as in *H. influenzae*.

Table 5. Regulon core and genes with potential taxon-specific regulation in the genomes of Vibrionales

Gene*	Annotation	VCH	VPA	VVU	PPR
VC0092	*lex*A, SOS regulator	+	+	+	+
VC0543	*rec*A, SOS coregulator	–	+	+	+
VC0852	*rec*N, DNA repair	+	+	+	+
VC1846	*ruv*A, Holliday junction DNA helicase	+	+	+	+
VC0190	*uvr*D, DNA helicase II	+	+	+	+
VC0394	*uvr*A, endonuclease ABC subunit A	+	+	+	+
VC0394	*uvr*B, endonuclease ABC subunit B	–	+	+	–
VC2043	*top*B, DNA topoisomerase III	+	+	+	+
VC2287	*din*P, DNA polymerase IV	+	+	+	+
VCA0291	*int*I4, site-specific recombinase	+	+	+	0
VC0081	Putative permease	+	+	+	+
VC0082	*yig*N, recombinase	+	+	+	+
VC0668	*mut*H, DNA mismatch repair protein	–	+	+	–
VC2711	*rec*G, ATP-dependent DNA helicase	+	+	–	–
VC0517	*rpo*D, RNA polymerase σ factor	–	+	+	–
VC1878	*msb*A, ATP-dependent transport protein	–	+	0	+

* The genes are designated as in *V. cholerae*.

Pasteurellales

LexA regulates *lex*A, *rec*N, *ruv*A, *rec*A, and *uvr*D in the *H. influenzae, H. ducreyi*, and *P. multocida* genomes (Table 4).

The LexA-binding sites upstream of *uvr*A and *ssb*, which have been experimentally assigned to the SOS system, are unstable. The site upstream of *uvr*A was found in the *H. influenzae* and *P. multocida* genomes and the site upstream of *ssb* was found in the *H. influenzae* and *H. ducreyi* genomes.

Potential taxon-specific regulation was assumed for *mfd* (potential transcription–repair coupling factor). A LexA-binding site was found upstream of *mfd* in two (*H. influenzae* and *P. multocida*) out of the three genomes.

The *E. coli mfd* gene is the closest homolog of Pasteurellales *mfd* with the function verified experimentally. When the template DNA strand is damaged during transcription, *E. coli* Mfd facilitates its repair by interacting with UvrA [43]. The upstream region of *mfd* lacks a potential regulatory site in the *E. coli* genome. Notwithstanding, the functional association of Mfd with the SOS system indicates that the LexA-dependent regulation of *mfd* is possible at least in some genomes.

Vibrionales

Among the core genes of the LexA regulon, *lex*A, *rec*N, *uvr*A, *uvr*D, *ruv*A, and *din*P displayed the most conserved regulation in the *V. cholerae, V. parahaemolyticus, V. vulnivicus*, and *P. profundum* genomes (Table 5).

All genomes examined have a signal upstream of *VC0081*, which codes for a potential weakly specific toxin exporter and belongs to COG0697 (Fig. 5). Since *VC0081* orghologs were not found in the *E. coli* genome, *VC0081* is a potential taxon-specific member of the SOS response.

In addition, all genomes have a potential signal upstream of *VC0082*, coding for recombinase (Fig. 6). *VC0082* is homologous to *E. coli yig*N, which is regulated by LexA. However, the distance between the potential regulatory site and the coding region is much the same in *VC0081* and *VC0082*, suggesting that the

```
VPA | VP0093    TAGACACTGGATAAATGTCCAGTTTGTTGGATGAAAAATC---
VVU | VV10911   TACATACTGGATAAATGTCCAGTTTGTTGCACAGAAAATCCTC
VCH | VC0081    TCGGCACTGGATAAATGTCCAGTTTGTTGCGCGTTTATTCTC-
PPR | PPR0114   TGGATACTGGATAAATGTCCAGCA-GTTATCCGTTAACGA---
                 *     ****************    ***         *
```

Fig. 5. Alignment of the *VC0081* upstream regions.

two genes belong to a divergon. Thus, it is unclear whether *VC0081* is actually regulated by LexA.

The signals upstream of *rec*G, *rpo*D, *mut*H, *uvr*B, and *int*I are conserved to a lesser extent. Potential regulatory sites upstream of these genes were found only in two genomes, *V. parahaemolyticus* and *V. vulnificus*. The *V. cholerae* and *P. profundum* orthologs of these genes lack LexA-binding sites.

Pseudomonadales and Alteromonadales

Since Pseudomonadales and Alteromonadales are phylogenetically close, their genomes were pooled in comparative genomic analysis. The *P. aeruginosa, P. putida, P. syringae*, and *S. oneidensis* genomes have regulatory signals upstream of *lex*A, *rec*A, and *rec*N, whose role in the SOS regulon is experimentally verified (Table 6).

The potential regulatory site upstream of *top*B, which codes for DNA polymerase III, was found in three genomes (*P. putida, P. syringae*, and *S. oneidensis*). The signal is conserved to a relatively low extent, one arm of the palindrome being conserved better than the other (Fig. 7).

DISCUSSION

In Enterobacteriales, we found six genes that are potentially regulated by LexA and have upstream signals only in the genomes of this group: *ycc*R (DNA-transforming protein), *ydj*Q (nuclease subunit of the exonuclease complex), *sbm*C (gyrase inhibitor), *ygj*F (G/U mismatch-specific DNA glycosylase), *ots*B (osmotic pressure regulator), *ydj*M (membrane-associated metal-dependent hydrolase), and *yfi*K (transport protein).

In Pasteurellales, only one gene potentially regulated by LexA had an upstream signal only in the genomes of the group. This was *mfd*, whose product is involved in repair and interacts with UvrA, participating in the SOS response.

In Vibrionales, we found two genes with potential taxon-specific regulation: *VC0081*, which codes for a weakly specific toxin transporter, and *VC0082*, which codes for recombinase.

```
VVU | VV10910 GCAACAA--ACTGGACATTTATCCAGTATGTAAA
VPA | VP0094   CCAACAA--ACTGGACATTTATCCAGTGTCTAAA
VCH | VC0082   GCAACAA--ACTGGACATTTATCCAGTGCCGAAA
PPR | PPR0115  CGGATAACTGCTGGACATTTATCCAGTATCCACT
                  *  **    ****************     *
```

Fig. 6. Alignment of the *VC0082* upstream regions.

In the pooled sample of Pseudomonadales and Alteromonanales, potential taxon-specific regulation by LexA was detected for *top*B, coding for DNA polymerase III. This enzyme has not been implicated in SOS-response DNA repair as of yet.

Thus, genes potentially regulated by LexA and having a signal only in the genomes of one group (suggesting taxon-specific regulation) were found in all groups examined (Enterobacteriales, Pasteurellales, Vibrionales, Pseudomonadales, and Alteromonadales). This finding indicates that the evolution of the SOS system differed among the phylogenetic groups of bacteria, depending on the environmental conditions and other group-specific factors.

One of the filters used to select the potential genes of the SOS regulon was that the function of a candidate gene corresponded to the cell reactions involved in the SOS response. The genes selected with this filter are implicated in DNA synthesis, which is essential for DNA repair; xenobiotic transport; and the maintenance of homeostasis.

In some cases, the upstream LexA-binding site was rather unstable but the gene function was associated with the SOS response. Several genes had a potential upstream regulatory site only in one of the genomes examined. The LexA-dependent regulation of these genes cannot be verified by purely bioinformatics methods. Since their possible role in the SOS response deserves experimental verification, the genes are listed below.

The *H. ducreyi* genome has a potential regulatory signal upstream of *rad*A, coding for a DNA repair protein, and *HD0897*, coding for integrase/recombinase.

The *V. parahaemolyticus* genome has a signal upstream of *VP2379*, which belongs to COG3141 (uncharacterized B-cell receptor). This cluster

Table 6. Regulon core and genes with potential taxon-specific regulation in the genomes of Pseudomonadales and Alteromonadales

Gene*	Annotation	PAE	PPU	PST	SON
*lex*A	*lex*A, SOS regulator	+	+	+	+
*rec*A	*rec*A, SOS coregulator	+	+	+	+
*rec*N	*rec*N, DNA repair	+	+	+	+
*top*B	*top*B, DNA topoisomerase III	0	+	+	+

* The genes are designated as in *P. aeruginosae*.

```
PPU | topB  ACATCCCCGCTGACAAAAA**ACCTGTACATCCATCCAG-AT**AAAACTTGCCTCTGCGC
PST | topB  ACAGCGCCAAAACGCCTTT**ATCTGTATATCCATACAG-AT**AAAACTTGCCCCT-CGT
SON | topB  CCTGCTTTGGGCTAAATTA**TACTGTAATTATATCCAGTA**TTTAGTGTAGATATGAT-
             *  *                *****  *  ** *** **  *   *     *
```

Fig. 7. Alignment of the *top*B 5'-noncoding regions.

includes *E. coli yeb*G, whose potential regulation by LexA has been assumed previously (our unpublished data).

In *V. parahaemolyticus*, a putative LexA-binding site was also found upstream of *VP2449*, coding for a weakly specific toxin exporter. This gene belongs to COG0534. The cluster includes *E. coli din*F, which is responsible for xenobiotic resistance and, on experimental evidence, is regulated by LexA and plays a role in the SOS response.

The *V. vulnificus* genome has regulatory sites upstream of the genes that belong to a transposon and code for transposase (*VV12451, VV12456, VV12476, VV12517, VV12529, VV12539*, and *VV12548*) and integrase (*VV12401*). SOS regulation is important for mobile elements because host DNA integrity is essential for their normal function. Thus, it seems logical that the genes responsible for the excision and integration of a mobile element have upstream regulatory motifs.

The *P. putida* genome has a signal upstream of *PP4068*, which codes for a transcriptional regulator of the Cro/CI family. The regulator is homologous to the phage λ CI repressor of *E. coli* (Fig. 4). Since the SOS regulation of the *E. coli* CI repressor is well known, it is likely that *P. putida PP4068* is also regulated by LexA.

In addition, the *P. putida* genome has a potential LexA-binding site upstream of *grp*E, coding for a heat shock protein. The protein is possibly involved in the SOS response when DNA damage is caused by an extreme temperature.

Interestingly, *lex*A is duplicated in the *P. putida* and *P. syringae* genomes. As phylogenetic analysis revealed, *P. putida lex*A-2 is orthologous to *P. syringae lex*A-2 while *P. putida lex*A-1 is orthologous to *P. syringae lex*A-1. The *P. aeruginosae* genome harbors only one *lex*A copy, which displays a higher homology to *P. syringae lex*A-2 and *P. putida lex*A-1 than to the paralogous gene pair. Putative LexA-binding sites were found upstream of *P. aeruginosa lex*A, *P. syringae lex*A-2, and *P. putida lex*A-1. It is possible that these genes do code for functional regulators, while their paralogs (*P. putida lex*A-2 and *P. syringae lex*A-1) have lost this function and no longer code for the SOS repressor LexA.

ACKNOWLEDGMENTS

This work was supported by the Howard Hughes Medical Institute (grant no. 55005610), INTAS (grant no. 05-1000008-8028), and the program Molecular and Cell Biology of the Russian Academy of Sciences.

REFERENCES

1. Walker C.G. 1996. The SOS response of *Escherichia coli*. In: *Escherichia coli* and *Salmonella*: *Cellular and Molecular Biology.* Neidhardt F.C., et al., Washington, DC: ASM Press, pp. 1400–1416.
2. Defais M., Fanquet P., Radman M., Errera M. 1971. Ultraviolet reactivation and ultraviolet mutagenesis of λ in different genetic systems. *Virology.* **43**, pp. 495–503.
3. Radman M. 1974. Phenomenology of an inducible mutagenic DNA repair pathway in *Escherichia coli*: SOS repair hypothesis. In: *Molecular and Environmental Aspects of Mutagenesis*. Eds. Prakash L. et al., Springfield, IL: Charles C. Thomas, pp. 128–142.
4. Aertsen A., van Houdt R., Vanoirbeek K., Michiels C.W. 2004. An SOS response induced by high pressure in *Escherichia coli. J. Bacteriol*. **186**, 6133–6141.
5. Qiu X., Sundin G.W., Wu L., Zhou J., Tiedje J.M. 2005. Comparative analysis of differentially expressed genes in *Shewanella oneidensis* MR-1 following exposure to UVC, UVB, and UVA radiation. *J. Bacteriol*. **187**, 3556–3564.
6. Michel B. 2005. After 30 years of study, the bacterial SOS response still surprises us. *PloS Biol*. **33**, 1174–1176.
7. Little J.W., Mount D.W., Yanisch-Perron C.R. 1981. Purified LexA protein is a repressor of the *rec*A and *lex*A genes. *Proc. Natl. Acad. Sci. USA*. **78**, 4199–4203.
8. Nuyts S., van Mellaert L., Barbe S., et al. 2001. Insertion or deletion of the Cheo box modifies radiation inducibility of *Clostridium* promoters. *Appl. Environ. Microbiol*. **67**, 4464–4470.
9. Brent R., Ptashne M. 1981. Mechanism of action of the *lex*A gene product. *Proc. Natl. Acad. Sci. USA*. **78**, 4204–4208.
10. Galhardo R.S., Rocha R.P., Marques M.V., Menck C.F.M. 2005. An SOS-regulated operon involved in damage-inducible mutagenesis in *Caulobacter crescentus. Nucleic Acids Res*. **33**, 2603–2614.
11. Winterling K.W., Levine A.S., Yasbin R.E., Woodgate R. 1997. Characterization of DinR, the *Bacillus subtilis* SOS repressor. *J. Bacteriol*. **179**, 1698–1703.
12. Winterling K.W., Chafin D., Hayes J.J., Sun J., Levine A.S., Yasbin R.E., Woodgate R. 1998. The *Bacillus subtilis* DinR binding site: Redefenition of the consensus sequence. *J. Bacteriol*. **180**, 2201–2211.

13. Erill I., Jara M., Salvador N., Escribano M., Campoy S., Barbe J. 2004. Differences in LexA regulon structure among Proteobacteria through in vivo assisted comparative genomics. *Nucleic Acids Res*. **32**, 6617–6626.
14. Mazon G., Erill I., Campoy S., Cortes P., Forano E., Barbe J. 2004. Reconstruction of the evolutionary history of the LexA-binding sequence. *Microbiology.* **150**, 3783–3795.
15. Stormo G.D., Hartzell G.W. 1989. Identifying protein-binding sites from unaligned DNA fragments. *Proc. Natl. Acad. Sci. USA*. **86**, 1183–1187.
16. Lawrence C.E., Reilly A.A. 1990. An EM algorithm for the identification and characterization of common sites in unaligned biopolymers sequence. *Proteins*. **7**, 41–51.
17. Van Helden J., Andre B., Collado-Vides J. 1998. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J. Mol. Biol*. **281**, 827–842.
18. Lawrence C.E., Altschul S.F., Boguski M.S., Liu J.S., Neuwald A.F., Wooton J.C. 1993. Detecting subtle sequence signals: A Gibbs sampling strategy for multiple alignment. *Science*. **262**, 208–214.
19. De Henestrosa A.F., Ogi T., Aoyagi S., Chafin D., Hayes J.J., Ohmori H., Woodgate R. 2000. Identification of additional genes belonging to the LexA regulon in *Escherichia coli. Mol. Microbiol*. **35**, 1560–1572.
20. Erill I., Escribano M., Campoy S., Barbe J. 2003. In silico analysis reveals substantial variability in the gene contents of the γ proteobacteria LexA regulon. *Bioinformatics*. **19**, 2225–2236.
21. Resendis-Antonio O., Freyre-Gonzalez J.A., Menchaca-Mendez R., Gutierrez-Rios R.M., Martinez-Antonio A., Avila-Sanchez C., Collado-Vides J. 2005. Modular analysis of the transcriptional regulatory network of *E. coli. Genetics*. **21**, 16–20.
22. Gelfand M.S., Mironov A.A. 1999. Computer analysis of regulatory signals in complete bacterial genomes. *Mol. Bio*. **33**, 772–778.
23. Mironov A.A., Vinokurova N.P., Gelfand M.S. 2000. Software for analysis of bacterial genomes. *Mol. Biol*. **34**, 253–262.
24. Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. **25**, 4876–4882.
25. Felsenstein J. 1989. PHYLIP-Phylogeny Inference Package (Version 3 .2). *Cladistics*. **5**, 164–166.
26. http://www.ncbi.nlm.nih.gov/Genbank/index.html
27. Blattner F.R., Plunkett G., Bloch C.A., et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science*. **277**, 1453–1474.
28. http://genome.wustl.edu/gsc/
29. McClelland M., Sanderson K.E., Spieth J., et al. 2001. Complete genome sequence of *Salmonella enterica* serovar *typhimurium* LT2. *Nature*. **413**, 852–856.
30. Wei J., Goldberg M.B., Burland V., et al. 2003. Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect. Immunol*. **71**, 2775–2786.
31. http://www.sanger.ac.uk/Projects/Microbes/
32. Duchaud E., Rusniok C., Frangeul L., et al. 2003. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens. Nature Biotechnol*. **21**, 1307–1313.
33. http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi@val=AE017143
34. Cole S.T., Eiglmeier K., Parkhill J., et al. 2001. Massive gene decay in the leprosy bacillus. *Nature*. **6823**, 1007–1011.
35. May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V. 2001. Complete genomic sequence of *Pasteurella multocida* PM70. *Proc. Natl. Acad. Sci. USA*. **98**, 3460–3465.
36. Makino K., Oshima K., Kurokawa K., et al. 2003. Genome sequence of *Vibrio parahaemolyticus*: A pathogenic mechanism distinct from that of *V. cholerae. Lancet*. **361**, 743–749.
37. Kim Y.R., Lee S.E., Kim C.M., et al. 2003. Characterization and pathogenic significance of *Vibrio vulnificus*. Antigens preferentially expressed in septicemic patients. *Infect. Immunol*. **71**, 5461–5471.
38. http://www.tigr.org/cgi-bin/
39. Heidelberg J., Paulsen I., Nelson K., et al. 2002. Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis. Nat. Biotechnol*. **20**, 1118–1123.
40. http://tcdb.ucsd.edu/
41. Altschul S.F., Madden T. L., Schlffer A. A., Zhang J., Zhang Z., Miller W., Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res*. **25**, 3389–3402.
42. Tatusov R.L., Koonin E.V., Lipman D.J. 1997. A genomic perspective on protein families. *Science*. **278**, 631–637.
43. Selby C.P., Sancar A. 1995. Structure and function of transcription–repair coupling factor: 1. Structural domains and binding properties. *J. Biol. Chem*. **270**, 4882–4889.