An SOS Regulon under Control of a Noncanonical LexA-Binding Motif in the Betaproteobacteria

Neus Sanchez-Alberola, a,b Susana Campoy, a David Emerson, c Jordi Barbé, a Ivan Erill b
Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Barcelona, Spain; Department of Biological Sciences, University of Maryland Baltimore County, Baltimore, Maryland, USA; Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, USA

ABSTRACT
The SOS response is a transcriptional regulatory network governed by the LexA repressor that activates in response to DNA damage. In the Betaproteobacteria, LexA is known to target a palindromic sequence with the consensus sequence CTGT-N8-ACAG. We report the characterization of a LexA regulon in the iron-oxidizing betaproteobacterium Sideroxydans lithotrophicus. In silico and in vitro analyses show that LexA targets six genes by recognizing a binding motif with the consensus sequence GAACGaaCGTTC, which is strongly reminiscent of the Bacillus subtilis LexA-binding motif. We confirm that the closely related Gallionella capsiferriformans shares the same LexA-binding motif, and in silico analyses indicate that this motif is also conserved in the Nitrosomonadales and the Methylophilales. Phylogenetic analysis of LexA and the alpha subunit of DNA polymerase III (DnaE) reveal that the organisms harboring this noncanonical LexA form a compact taxonomic cluster within the Betaproteobacteria. However, their lexA gene is unrelated to the standard Betaproteobacteria lexA, and there is evidence of its spread through lateral gene transfer. In contrast to other reported cases of noncanonical LexA-binding motifs, the regulon of S. lithotrophicus is comparable in size and function to that of many other Betaproteobacteria, suggesting that a convergent SOS regulon has reevolved under the control of a new LexA protein. Analysis of the DNA-binding domain of S. lithotrophicus LexA reveals little sequence similarity with that of other LexA proteins targeting similar binding motifs, suggesting that network structure may limit site evolution or that structural constrains make the B. subtilis-type motif an optimal interface for multiple LexA sequences.

IMPORTANCE
Understanding the evolution of transcriptional systems enables us to address important questions in microbiology, such as the emergence and transfer potential of different regulatory systems to regulate virulence or mediate responses to stress. The results reported here constitute the first characterization of a noncanonical LexA protein regulating a standard SOS regulon. This is significant because it illustrates how a complex transcriptional program can be put under the control of a novel transcriptional regulator. Our results also reveal a substantial degree of plasticity in the LexA recognition domain, raising intriguing questions about the space of protein-DNA interfaces and the specific evolutionary constrains faced by these elements.

In order to survive, bacteria react to transient environmental insults through the activation of coordinated transcriptional programs, collectively known as stress responses. Among these, the SOS response to DNA damage is one of the most widely conserved and well characterized (1–3). In Escherichia coli, where it was originally described, the SOS response controls the expression of almost 40 genes involved in DNA repair, translesion synthesis, and cell division inhibition (2, 4). The SOS response is coordinated by the transcriptional repressor LexA, which binds to specific sites upstream of regulated operons and blocks transcription initiation (1, 2). DNA damage is sensed by the RecA protein, which forms active nucleoprotein filaments upon binding single-stranded DNA fragments formed at stalled replication forks (5). Activated RecA promotes self-cleavage of the LexA repressor, leading to the induction of the SOS response, which includes the lexA and recA genes (2, 4, 6).

Most major bacterial clades possess a functional lexA gene whose product controls an SOS regulon of variable size, with notable exceptions, including the Bacteroidetes, the Epsilonproteobacteria or the Streptococcales (3, 7, 8). In contrast to other broadly conserved transcriptional regulators, the DNA-binding motif of LexA has undergone several drastic changes in the course of evolution (3). For instance, LexA targets a 16-bp palindromic binding motif with a consensus sequence (CTGT-N8-ACAG) in E. coli, a 12-bp palindrome in Bacillus subtilis (GAAC-N4-GTTC), and a 15-bp direct repeat in Caulobacter vibrioides (GTTC-N7-GTTC) (9–11). These motifs are known to be monophyletic in their respective major clades (Gamma proteobacteria, Firmicutes, and Alphaproteobacteria) (12–14), but the description of three distinct LexA-binding motifs in the Deltaproteobacteria bears testament to the substantial recognition plasticity of the LexA repressor over different evolutionary scales. In spite of these changes, the LexA-binding motif, the SOS response maintains a consistent regulatory

Received 21 January 2015 Accepted 9 May 2015
Accepted manuscript posted online 18 May 2015
Editor: R. L. Gourse
Address correspondence to Ivan Erill, erill@umbc.edu.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/JB.00035-15.
Copyright © 2015, American Society for Microbiology. All Rights Reserved. doi:10.1128/JB.00035-15
function across bacterial groups, providing a unique perspective into the coevolution of transcriptional regulators and their regulatory networks (3, 4, 9, 11).

In the Betaproteobacteria, in vivo and in silico studies have established that the SOS response is regulated by a LexA protein targeting a binding motif closely related to that observed in E. coli and other Gammaproteobacteria (12, 15, 16). Comparative genomic and transcriptomic data, however, indicate that the SOS regulon of the Betaproteobacteria displays some fundamental differences from that reported for many Gammaproteobacteria. For instance, the Betaproteobacteria LexA protein does not regulate the recN gene or the recombination repair operon ruvAB. Beyond the core SOS genes shared with the Gammaproteobacteria (lexA, recA, and uvrA), the Betaproteobacteria SOS response regulates a gene coding for a predicted photolyase (spILb) and a mutagenesis cassette (imuAB-dnAE2) also regulated by LexA in many bacterial groups (15, 17). Sideroxydans lithotrophicus and Gallionella capsiferriformans are chemolithoautotrophic Betaproteobacteria capable of using iron as their main electron source (18). The complete genome sequences of these two representatives of the Gallionellales order reveal the presence of a lexA gene with low sequence identity to its well-characterized homologs in other Betaproteobacteria orders (19). Here, we combine in silico and in vitro techniques to characterize the LexA-binding motif and SOS regulon of these two bacterial species. Our results confirm that the lexA gene of the Gallionellales and other related bacteria is not closely related to the canonical Betaproteobacteria lexA and show that its product targets a binding motif resembling that recognized by B. subtilis LexA. Furthermore, the composition of the SOS regulon in S. lithotrophicus suggests that the SOS regulatory network under the control of this novel lexA arose through a process of convergent evolution.

MATERIALS AND METHODS

Genome, motif, and sequence data. Available complete genome sequences for the species under study were downloaded from the NCBI RefSeq database in GenBank (.gbk) format (20). Experimentally validated transcription factor-binding motifs for individual species and clade aggregates were downloaded from the CollecTF database (21). Orthologs for S. lithotrophicus and G. capsiferriformans LexA on complete genome sequences were identified with PSI-BLAST and validated as best reciprocal BLAST hits using BLASTP with E values lower than 10−20 (14, 22). LexA and DnaE orthologs from complete genome sequences were compiled based on their annotation in RefSeq and BLAST searches and were then validated through reciprocal BLAST. LexA and DnaE orthologs were compiled to provide broad coverage of selected phyla and classes, while maintaining a relatively low total number of taxa for phylogenetic inference. For comparative genomics, orthologs of putatively regulated genes were determined as best reciprocal BLAST hits of their encoded protein sequences. Upstream regions (−250, +50) of lexA orthologs were obtained from the NCBI GenBank database using custom Python scripts. Circular genome maps for S. lithotrophicus and G. capsiferriformans were generated using DNAPlotter (23) with default parameters. Sequence identity from pairwise alignments and molecular weights for LexA proteins were computed using the BioWord add-on for Microsoft Word (24).

Site search, motif discovery, and comparative genomics. Whole-genome sequences were scanned for inferred and experimentally validated motifs using xFITOM with the sequence information content (R) scoring method and a motif-normalized threshold as reported previously (25–27). Putative LexA-binding sites were associated with annotated genes if they were located within bp −250 to +50 from the predicted translation start point for the gene and their score was less than one standard deviation below the average score of the original collection (13, 15). For the comparative genomics analysis of the Gallionellales taxonomic cluster, genes were defined as LexA-regulated if candidate sites were located in at least two species of the cluster. The LexA-binding motif of the Gallionellales cluster was identified using MEME on the upstream (−250, +50) regions of identified lexA gene orthologs, searching for palindromic 10- to 30-bp motifs with N instances per sequence and with otherwise default parameters (28). Putative −35 and −10 promoter elements were identified with the phISITE PromoterHunter web service and the PePPER tool-box with default parameters (29, 30).

Multiple sequence alignments and phylogenetic inference. Multiple sequence alignments of DNA and protein sequences were performed using CLUSTAL W (31). For LexA and DnaE sequence alignments, the information of the E. coli LexA and DnaE crystal structures (32, 33) was used in the UniProt records POA7C2 and P10443 to define penalty masks in profile alignment mode. Editing and graphical representation of alignments were performed with BioEdit (34). Sequence logos were created using aligned DNA or protein sequences with the WebLogo service (35). For phylogenetic inference, multiple sequence alignments of LexA and DnaE protein sequences were generated with CLUSTAL W using variable (5, 10, and 25) gap opening penalties for the multiple-alignment stage (17). These alignments were integrated with a local LALIGN alignment for optimization using T-COFFEE (36), and the resulting alignment was processed with GBLOCKS with the half-gaps setting and otherwise default parameters to select conserved positions for reliable phylogenetic inference (37). As the closest relatives to Gammaproteobacteria and Betaproteobacteria, Alphaproteobacteria species were included as the outgroup for DnaE tree inference (38). Alphaproteobacteria LexA protein sequences are known to present a 33-amino-acid insertion that disrupts the multiple-sequence alignment (39). To enhance alignment quality and phylogenetic resolution, Gram-positive LexA protein sequences were used as the outgroup for LexA tree inference. Bayesian phylogenetic inference on T-COFFEE-optimized alignments was performed with MrBayes (40). Twelve Metropolis-Coupled Markov Chain Monte Carlo runs with four independent chains were carried out for 500,000 generations, and the resulting consensus tree was plotted with TreeView (41).

EMSA. Electrophoresis mobility shift assays (EMSA) were performed to validate computationally predicted LexA-binding sites in the promoter region of putative G. capsiferriformans and S. lithotrophicus SOS genes. G. capsiferriformans ES-2 DNA was extracted from phosphate buffered (50 mM) saline (pH 8.0)-washed pellets containing cells and iron oxides using a MoBio PowerSoil device (MoBio, Carlsbad, CA) DNA isolation kit (42). The G. capsiferriformans ES-2 lexA gene was amplified using suitable primers (see Table S1 in the supplemental material) and cloned into a pET15b vector (see Table S2 in the supplemental material). Overexpression and purification of the G. capsiferriformans ES-2 and E. coli and B. subtilis LexA proteins was performed as described previously for other LexA proteins (26, 43, 44). DNA probes were constructed using two complementary 61-bp synthetic oligonucleotides centered on the LexA-binding sites of interest (see Table S1 in the supplemental material). Overexpression and purification of the G. capsiferriformans ES-2 and E. coli and B. subtilis LexA proteins was performed as described previously for other LexA proteins (26, 43, 44). DNA probes were constructed using two complementary 61-bp synthetic oligonucleotides centered on the LexA-binding sites of interest (see Table S1 in the supplemental material). Overexpression and purification of the G. capsiferriformans ES-2 and E. coli and B. subtilis LexA proteins was performed as described previously for other LexA proteins (26, 43, 44). DNA probes were constructed using two complementary 61-bp synthetic oligonucleotides centered on the LexA-binding sites of interest (see Table S1 in the supplemental material). Overexpression and purification of the G. capsiferriformans ES-2 and E. coli and B. subtilis LexA proteins was performed as described previously for other LexA proteins (26, 43, 44). DNA probes were constructed using two complementary 61-bp synthetic oligonucleotides centered on the LexA-binding sites of interest (see Table S1 in the supplemental material). Overexpression and purification of the G. capsiferriformans ES-2 and E. coli and B. subtilis LexA proteins was performed as described previously for other LexA proteins (26, 43, 44). DNA probes were constructed using two complementary 61-bp synthetic oligonucleotides centered on the LexA-binding sites of interest (see Table S1 in the supplemental material).
RESULTS AND DISCUSSION

In silico analysis of the lexA promoter of S. lithotrophicus and G. capsiferriformans. The complete genome sequences of both S. lithotrophicus and G. capsiferriformans contain a gene (Slit_2300 and Galf_0635, respectively) with a predicted product annotated as LexA repressor. The two predicted lexA genes are in opposite ends of the genome (see Fig. S1 in the supplemental material) and, following the overall pattern reported for these two genomes, the genomic surroundings of these lexA genes reveal no evidence of synteny (19). Multiple-sequence alignment of the amino acid sequence of these predicted LexA proteins shows that they are homologous, sharing 78% sequence identity with each other, but only 24% sequence identity versus the LexA protein of E. coli (coded by b4043) or that of the more closely related Burkholderia thailandensis (coded by BTH_I2481). The latter figure is comparable to the sequence identity (28%) that S. lithotrophicus and G. capsiferriformans LexA proteins share with their ortholog (coded by BSU17850) in the more distantly related B. subtilis (see Fig. S2 in the supplemental material). An in silico analysis of the S. lithotrophicus and G. capsiferriformans genome sequences using the LexA-binding motif targeted by E. coli and B. thailandensis LexA did not yield any significant hits in the promoter region of known SOS genes, including lexA. Sequence alignment of the intergenic region upstream of these two lexA genes revealed the presence of two conserved imperfect palindromes in a tandem configuration and overlapping the region encompassing the predicted −35 and −10 promoter elements (Fig. 1). Except for one reported case (45), the lexA gene is typically regulated by its functional product through binding at specific target sites on its upstream sequence, and this property has been effectively leveraged to infer the binding motif of LexA and other transcriptional repressors through comparative genomics (14, 46). Tandem configurations resembling the conserved palindromes shown in Fig. 1 have been reported previously for functional LexA-binding sites upstream of lexA and other SOS regulated genes (9, 12, 14, 47, 48). Furthermore, the identified palindromes display a consensus sequence (GAAGC-N2-GTTC) that is strongly reminiscent of the B. subtilis LexA-binding motif. Functional LexA-binding motifs bearing strong similarity to the B. subtilis one have been reported in Firmicutes, Actinobacteria, Cyanobacteria, and chloroflexi and for a LexA paralog in Gammaproteobacteria (26, 48, 49–54). Taken together, these data strongly suggested that the identified tandem palindromes were functional LexA-binding sites.

Characterization of the SOS regulon in the Gallionellales. To study the composition of the SOS regulatory network under the control of the Gallionellales LexA protein, we performed an in silico genome-wide search for putative LexA-binding sites using the four identified conserved palindromes as the putative LexA-binding motif. The results shown in Table 1, reveal both similarities and differences in the composition of the putative S. lithotrophicus and G. capsiferriformans SOS response. The inferred LexA regulons comprise five shared genes coding for essential SOS proteins: the LexA repressor, the recombination proteins RecA and RecX, and the umuC-encoded DNA polymerase V. In G. capsiferriformans, the predicted LexA regulon encompasses two copies of DNA polymerase V, since its lexA gene (Galf_0635) forms a putative operon with umuD (Galf_0634–Galf_0635). These results hence point toward a core LexA regulon devoted to transcription synthesis and are in broad agreement with previous reports on the composition of the SOS response in multiple bacterial clades (3). In addition to these core genes, the LexA protein of S. lithotrophicus appears to regulate three additional genes: a spIB

![FIG 1 Sequence alignment of the region immediately upstream of the S. lithotrophicus (Slit_2300) and G. capsiferriformans (Galf_0635) lexA genes. Predicted −35 and −10 promoter elements are boxed. The predicted translational start site is underlined. The conserved palindromic sequences corresponding to putative LexA-binding sites are shaded in gray.](http://jb.asm.org/)

**TABLE 1** Putative LexA-binding sites identified through a genome-wide in silico search.

<table>
<thead>
<tr>
<th>Score</th>
<th>Site</th>
<th>Strand</th>
<th>d_{TLS}</th>
<th>d_{−10}</th>
<th>Position</th>
<th>Locus</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.00</td>
<td>AAGAACGTCTGTCAGTCTG</td>
<td>+</td>
<td>−53</td>
<td>−16</td>
<td>671728</td>
<td>Galf_0635</td>
<td>LexA</td>
</tr>
<tr>
<td>17.41</td>
<td>GAGAACGTCTGTCAGTCTG</td>
<td>+</td>
<td>−50</td>
<td>−16</td>
<td>2589520</td>
<td>Galf_2402</td>
<td>UmuDC</td>
</tr>
<tr>
<td>17.41</td>
<td>GAGAACGTCTGTCAGTCTG</td>
<td>+</td>
<td>−28</td>
<td>+9</td>
<td>671719</td>
<td>Galf_0635</td>
<td>LexA</td>
</tr>
<tr>
<td>17.00</td>
<td>AAGAACGTCTGTCAGTCTG</td>
<td>+</td>
<td>−72</td>
<td>−37</td>
<td>354879</td>
<td>Galf_0508</td>
<td>RecA</td>
</tr>
<tr>
<td>18.77</td>
<td>AAGAACGTCTGTCAGTCTG</td>
<td>+</td>
<td>−38</td>
<td>+8</td>
<td>819364</td>
<td>Slit_0831</td>
<td>UmuDC</td>
</tr>
<tr>
<td>18.77</td>
<td>GAGAACGTCTGTCAGTCTG</td>
<td>+</td>
<td>−60</td>
<td>−16</td>
<td>2456160</td>
<td>Slit_2431</td>
<td>LexA</td>
</tr>
<tr>
<td>18.77</td>
<td>AAGAACGTCTGTCAGTCTG</td>
<td>+</td>
<td>−29</td>
<td>+8</td>
<td>2329266</td>
<td>Slit_2300</td>
<td>RecA</td>
</tr>
<tr>
<td>18.77</td>
<td>AAGAACGTCTGTCAGTCTG</td>
<td>+</td>
<td>−29</td>
<td>+3</td>
<td>819355</td>
<td>Slit_0831</td>
<td>UmuDC</td>
</tr>
<tr>
<td>18.77</td>
<td>GAGAACGTCTGTCAGTCTG</td>
<td>+</td>
<td>−57</td>
<td>+11</td>
<td>723063</td>
<td>Slit_0718</td>
<td>SpIB</td>
</tr>
<tr>
<td>18.77</td>
<td>GAGAACGTCTGTCAGTCTG</td>
<td>+</td>
<td>−60</td>
<td>−16</td>
<td>2456160</td>
<td>Slit_2431</td>
<td>(hyp)</td>
</tr>
<tr>
<td>18.18</td>
<td>GAGAACGTCTGTCAGTCTG</td>
<td>+</td>
<td>−60</td>
<td>−37</td>
<td>2133987</td>
<td>Slit_2105</td>
<td>RecA</td>
</tr>
<tr>
<td>17.18</td>
<td>GAGAACGTCTGTCAGTCTG</td>
<td>+</td>
<td>−76</td>
<td>−16</td>
<td>2329290</td>
<td>Slit_2300</td>
<td>LexA</td>
</tr>
</tbody>
</table>

a Site scores are reported in bits. Positive and negative symbols denote the strand on which sites were detected with a maximal score with respect to the leading strand in the annotated genome. The d_{TLS} and d_{−10} columns indicate the distance from the 5′ end of the site to the 5′ end of the predicted translation start site (TLS) and the −10 promoter element, respectively. The absolute genome position of the site is also given, as well as the NCBI locus tag for the gene with the closest TLS.

b For Galf_0508, two alternative −10 elements were predicted (see Fig. 55 in the supplemental material). The distances of the putative LexA-binding site to these elements are, respectively, −37 and +12 bp.
Betaproteobacterial Noncanonical LexA-Binding Motif

homolog encoding a predicted photolyase (Slit_0718), a hypothetical protein (Slit_2431), and a gene encoding a predicted filamentation protein containing a Fic domain (Slit_2431).

The in silico search results presented in Table 1 are congruent with LexA defining a canonical SOS response by recognizing B. subtilis-type LexA-binding sites in the promoter region of SOS genes. To validate that this putative LexA-binding motif was involved in the recognition of specific promoters by LexA, we performed EMSAs with purified LexA protein from G. capsiferriformans, E. coli, and B. subtilis on the promoter region of putative SOS genes from G. capsiferriformans and S. lithotrophicus. The EMSA results shown in Fig. 2 reveal that G. capsiferriformans LexA is able to bind targets identified in silico on both organisms, indicating that the LexA-binding motif of the Gallionellales is a 14-bp palindrome (AGAACG-N2-CGTTCT) that closely resembles the B. subtilis motif. Titration experiments with competing DNA show the binding of G. capsiferriformans LexA binds specifically to the Gallionellales promoters (see Fig. S3 in the supplemental material). Further evidence for the association between the Gallionellales and Gram-positive LexA-binding motifs is provided by the fact that purified B. subtilis LexA binds canonical SOS targets (lexA and recA promoters) of both organisms. In contrast, E. coli LexA, which targets a LexA-binding motif very similar to that reported in the Betaproteobacteria B. thailandensis, Bordetella pertussis, and Ralstonia solanacearum (12, 16, 55), does not bind canonical target promoters in the Gallionellales. In spite of the overall similarity of their binding motifs, B. subtilis and G. capsiferriformans LexA proteins show distinct binding specificities (see Fig. S4 in the supplemental material). As suggested by its strong conservation in the identified G. capsiferriformans and S. lithotrophicus LexA-binding sites (Table 1), the spacer and terminal positions of the AGAACG-N2-CGTTCT motif play an important role in site recognition by G. capsiferriformans LexA. Together with the in silico analyses, the binding assays reported in Fig. 2, Fig. S3, and Fig. S4 demonstrate that LexA recognizes a noncanonical binding motif in the Gallionellales, and indicate that these noncanonical sites define a small but prototypical SOS network in S. lithotrophicus and G. capsiferriformans. In this regard, it is worth noting that the proximal end of all of the LexA-binding sites reported in Table 1 is at the most 6 bp away from the predicted core promoter sequence and that in 78% of the analyzed sequences there is one LexA-binding site located within the predicted core promoter sequence (see Fig. S5 in the supplemental material). The location of predicted LexA-binding sites hence not only supports the notion that these are functional but also indicates that in all of these instances LexA most likely operates as a direct repressor, inhibiting transcriptional initiation by blocking access to the RNA polymerase binding site (10, 56, 57, 58).

Phylogenetic and comparative genomics analysis of the Gallionellales LexA. The identification of a LexA-binding protein targeting a noncanonical binding motif within the Betaproteobacteria can be most readily explained by two alternative evolutionary scenarios: lateral gene transfer (LGT) or duplication, followed by divergence and deletion of the canonical Betaproteobacteria lexA gene. Previous research has postulated the former as the likely mechanism for the presence of an Alphaproteobacteria-like LexA-binding motif in the Acidobacteria and diverging motifs in the Leptospirales genus and the latter as the originating process for the existence of an additional LexA protein targeting a Cyanobacteria-like motif in several Gammaproteobacteria (39, 51, 59). To elucidate the likely origin of the Gallionellales LexA protein, we performed phylogenetic inference on the amino acid sequence of the housekeeping α subunit of DNA polymerase III (DnaE) and the LexA protein (17). In agreement with previous work, the DnaE phylogenetic tree (Fig. 3) establishes the Gallionellales as a distinct clade within the Betaproteobacteria (19, 60). Furthermore, it confirms that this order is most closely related with the Nitrosomonadales and the Methylophilales and that this cluster delineates an early branching point in the Betaproteobacteria radiation (19, 60). The LexA tree also clusters the Gallionellales with the Nitrosomonadales and the Methylophilales but places them in a highly divergent group far removed from the main Betaproteobacteria clade. In addition to members of these orders, the cluster encompassing the Gallionellales LexA also includes LexA proteins from several unrelated Burkholderiales and from a Deltaproteobacteria (Syntrophus aciditrophicus). Phylogenetic analysis of a nonhousekeeping SOS protein, UmuC, is consistent with the results obtained for DnaE (see Fig. S6 in the supplemental material), indicating that the origin of the Gallionellales lexA is not linked with a broader rearrangement of SOS-related components.

The robust clustering of the Gallionellales with both related and unrelated bacteria in the LexA tree suggests that these LexA sequences share a origin and that, in accordance, their LexA protein might also target the same binding motif. Relying on the assumption that the lexA gene is regulated by its product (14), we used the MEME motif discovery algorithm to identify overrepresented motifs in the promoter region of lexA homologs from species clustering with the Gallionellales LexA. The most significant result

FIG 2 EMSAs were performed with 61-bp fragments of G. capsiferriformans and S. lithotrophicus DNA centered on putative LexA-binding sites predicted in silico. The numbers between brackets denote the distance of the targeted site to the predicted TLS, as reported in Table 1. Locus tags for putative SOS genes are reported. Lanes E and B designate assays with E. coli and B. subtilis LexA proteins, respectively. For all other lanes, G. capsiferriformans LexA protein was used. The “+” and “−” symbols denote, respectively, the presence or absence of LexA protein. The molecular masses of these LexA proteins are as follows: E. coli (22.35 kDa), B. subtilis (22.85 kDa), and G. capsiferriformans (21.97 kDa).
FIG 3 Consensus tree of DnaE (top) and LexA (bottom) protein sequences. Branch support values are provided as Bayesian posterior probabilities. LexA-binding motifs for different clades are illustrated using sequence logos made with inferred binding sites (Gallionellales cluster) or aggregate experimental LexA-binding data downloaded from the CollecTF database. Different clades are shaded in alternating shades of gray. The cluster containing the Gallionellales DnaE and LexA proteins is boxed and lightly shaded. The scale bar on each tree denotes expected substitutions per residue. To enhance alignment quality, the DnaE tree was rooted using two Alphaproteobacteria LexA sequences as outgroup and the LexA tree was rooted with four Gram-positive sequences as outgroup. A DnaE tree rooted with Gram-positive sequences is provided in Fig. S7 in the supplemental material for a direct comparison. The protein accession numbers and full species names for all represented protein sequences are available in Table S3 in the supplemental material.
returned by MEME is a LexA-binding motif encompassing the one experimentally validated here for the Gallionellales (AGAAC-NT-TTCT) (Fig. 3). This motif was identified in the promoter region of lexA in the Gallionellales, Nitrosonomadales, Methylotherma versatilis, Burkholderia phytofirmans, and S. aciditrophicus. In addition, a related motif (TGTAC-N4-GTACA) was identified in the Methylophilales, Methylobacterium flagellatus, and Methylovorans glucosotrophus (Fig. 3). To further verify that the identified motifs constituted the LexA-binding motif of this heterogeneous cluster, we used these motifs to perform a comparative genomics analysis of the predicted LexA regulon in the cluster of species encompassing the Gallionellales LexA. The comparative genomics results (Table 2) reveal that the shared LexA regulon in this cluster is limited to lexA, with only the two Gallionellales species showing evidence of recA regulation. In spite of this, the inferred LexA regulons of several species present significant similarities, such as the regulation of umuDC operons and of the splB gene encoding a predicted photolyase, providing further support for the in vivo functionality of the LexA-binding sites experimentally validated in this work.

Given the short and variable nature of the LexA protein sequence, phylogenetic inference methods cannot fully resolve the origin of the Gallionellales LexA. Nonetheless, the combined results of Fig. 3 and Table 2 do provide conclusive evidence of several facts regarding the evolutionary history of this protein. The presence in B. phytofirmans, Allicyclus denitrificans, and Ralstonia pickettii of another LexA protein clustering with the Betaproteobacteria LexA clearly establishes that the Gallionellales LexA does not derive directly from the canonical Betaproteobacteria LexA. Instead, the presence of two lexA copies in these genomes is suggestive of an early duplication scenario at the branching point between Gammaproteobacteria and Betaproteobacteria, followed by selective deletion of the canonical lexA gene (17). On the other hand, the presence of a Gallionellales-type LexA in a single Deltaproteobacteria indicates that this lexA has been involved in at least one lateral gene transfer event. Furthermore, the results also establish that the LexA proteins clustering with the Gallionellales LexA share a ancestor and target variables of a LexA-binding motif similar to that reported in B. subtilis. The small size of the putative SOS regulon in these Gallionellales cluster species, and specifically the loss of recA regulation by LexA, has been described before in association with rapidly diverging lexA genes, lexA duplications and putative LGT (26, 39, 59, 61–64). In this setting, the description in S. lithothrophicus of a more elaborate LexA regulon, including a signature gene of the canonical Betaproteobacteria SOS response (splB) points to a process of convergent evolution toward the regulation of this gene. This notion is reinforced by the consistent regulation of umuDC operons in Gallionellales cluster species, since their phylogeny is congruent with a stable genetic composition of SOS genes in the Betaproteobacteria (see Fig. S6 in the supplemental material). The hypothesis of convergent evolution toward the regulation of a prototypical Betaproteobacteria SOS response applies not only to scenarios involving LGT, where novel binding sites for the transferred LexA must necessarily be evolved on target genes, but also to scenarios involving gene duplication, followed by deletion of the canonical lexA. In the latter, the dramatic changes observed in the sequence of the noncanonical LexA protein and its binding motif are suggestive of relaxed or absent selection. This is difficult to reconcile with preservation of regulon composition, implying that sites for the noncanonical lexA on target genes such as splB must have evolved anew after the deletion of the canonical lexA gene (65).

**Comparative analysis of the LexA-binding domain.** LexA proteins recognizing B. subtilis-type motifs have been reported in the Actinobacteria, the Cyanoabacteria, the chloroflexi, and some Gammaproteobacteria (26, 48, 49–54). Even though phylogenetic inference cannot fully resolve the origin of the Gallionellales LexA protein, the fact that it targets a B. subtilis-like binding motif suggested that a detailed analysis of its DNA-binding domain might shed light onto its evolutionary origin. The crystal structure of E. coli LexA bound to DNA, as well as docked models of B. subtilis LexA and numerous biochemical and genetic studies, has clearly established that the α3 helix of the N-terminal helix-turn-helix motif is responsible for the majority of the specific contacts with the symmetric conserved elements of LexA-binding motifs in these two species (33, 66–69). Leveraging this structural information, we performed a multiple-sequence alignment of amino acid sequences for LexA proteins targeting similar half-sites and analyzed the alignment segments corresponding to the α3 helix.

The multiple-sequence alignment results summarized in Fig. 4 show that the α3 helix of the Gallionellales LexA is not directly related to the α3 helix of any other LexA targeting a conserved

---

**TABLE 2 Results of the comparative genomics analysis of the Gallionellales cluster**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Site</th>
<th>Strand</th>
<th>dTLS (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lexA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bphyl_3718</td>
<td>CGAAGACGTTTCG</td>
<td>+</td>
<td>–4</td>
</tr>
<tr>
<td>Galf_0635</td>
<td>AGAAGACGTTTCG</td>
<td>–</td>
<td>–66</td>
</tr>
<tr>
<td>Mla_1781</td>
<td>TGTACATTCGTA</td>
<td>–</td>
<td>–13</td>
</tr>
<tr>
<td>M301_2214</td>
<td>CGAAGACGTTTCG</td>
<td>–</td>
<td>–69</td>
</tr>
<tr>
<td>Msip43_2108</td>
<td>TGTACAGTTCTAC</td>
<td>–</td>
<td>–30</td>
</tr>
<tr>
<td>Nmul_A2596</td>
<td>AGAACAACTTGTC</td>
<td>–</td>
<td>–54</td>
</tr>
<tr>
<td>Slii_2300</td>
<td>AGAACAACTTGTC</td>
<td>+</td>
<td>–32</td>
</tr>
<tr>
<td>SYN_00359</td>
<td>AGAACAACTTGTC</td>
<td>–</td>
<td>–14</td>
</tr>
<tr>
<td>recA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galf_0508</td>
<td>AGAACGATCGTTTC</td>
<td>+</td>
<td>–71</td>
</tr>
<tr>
<td>Slii_2105</td>
<td>AGAACGATCGTTTC</td>
<td>–</td>
<td>–89</td>
</tr>
<tr>
<td>umuDC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galf_2402</td>
<td>AGAACGATCGTTTC</td>
<td>–</td>
<td>–63</td>
</tr>
<tr>
<td>Msip43_1483</td>
<td>TGTACAGTTCTAC</td>
<td>–</td>
<td>–15</td>
</tr>
<tr>
<td>Slii_0831</td>
<td>AGAACGATCGTTTC</td>
<td>+</td>
<td>–14</td>
</tr>
<tr>
<td>splB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mla_1759</td>
<td>TGTACAGTTCTAC</td>
<td>+</td>
<td>–8</td>
</tr>
<tr>
<td>Slii_0718</td>
<td>AGAACGATCGTTTC</td>
<td>–</td>
<td>–57</td>
</tr>
</tbody>
</table>

* Putative LexA-binding sites resulting from a genome-wide search are reported only if present upstream of at least two orthologs for a given gene. Positive and negative symbols denote the strand on which sites were detected with a maximal score with respect to the leading strand in the annotated genome. The dTLS column indicates the distance from the 5′ end of the site to the 5′ end of the predicted translation start site (TLS). The NCBI locus tag for the gene with the closest TLS is also given.

---

**Beta proteobacterial noncanonical LexA-binding motif**

Betaproteobacteria

---

**August 2015 Volume 197 Number 16 Journal of Bacteriology**

---

**2627**

---

**Downloaded from** https://jb.asm.org/ on June 29, 2017 by guest
GAAC half-site. Rather, the Gallionellales cluster LexA α3 helix displays a mosaic of features from other α3 helices, such as the conserved S3K2S3 pattern seen in the Alphaproteobacteria or the R10L11 motif observed in the Firmicutes and the Cyanobacteria. Most importantly, the Gallionellales cluster LexA data illustrate that a similar binding motif (GAAC-N4-GTTC) can be targeted by LexA proteins with seemingly unrelated α3 helix sequences. Although the α3 helices of Actinobacteria, Firmicutes, and chloroflexi display substantial sequence similarity, the highly divergent α3 helix sequences of the Xanthomonadaceae/Pseudomonadaceae, the Cyanobacteria, and the Gallionellales cluster target a structurally similar motif that has been shown be bound by several noncognate LexA proteins (50, 51). Recent work on the PhoQ-PhoP two-component signal transduction system has shown that there is a remarkable degree of sequence plasticity in the PhoQ-PhoP interface and that evolution has explored only a small fraction of the space of the PhoP-specific PhoQ motifs, due to both physiological and biochemical constraints on available mutational pathways (70–72). The results presented here suggest that the same is true for the LexA protein-DNA interface, which displays a varied palette of α3 helix and LexA-binding motif pairs. Importantly, however, the available data indicate that the coevolution of the LexA recognition domain and its binding motif may face more restrictions than those displayed by two-component systems (70, 71, 73). In particular, the abundance of highly divergent α3 helix sequences targeting a similar, B. subtilis-type LexA-binding motif yields two nonexclusive scenarios. On the one hand, it suggests that the need for coordinated changes in multiple binding sites might limit the evolution of the DNA component of the regulatory system. On the other hand, in the same way that the PhoQ-PhoP interface sequence distribution is highly nonuniform (72), the B. subtilis-type LexA-binding motif may provide a broadly accessible interface to the canonical LexA structure, enabling its interaction with a large number of α3 helix sequences and therefore substantially concentrating the evolutionary exploration of the α3 helix sequence space.

ACKNOWLEDGMENTS
This study was supported by UMBC Special Research Assistantship/Initiative Support (SRAIS) and U.S. National Science Foundation (MCB-1158056) awards to I.E. and by Spanish Ministry of Science and Innovation (BFU2011-23478) and Generalitat de Catalunya (2014SGR572) awards to J.B. N.S.-A. was funded by a Fundació Cellex fellowship.

REFERENCES
Erill Anzaldi

Pruitt Cornish

Wade Emerson Carver

Ulrich August 2015 Volume 197 Number 16 jb.asm.org


Tapias A, Fernandez S, Alonso JC, Barbe J. 2002. Rhodobacter sphaeroides LexA has dual activity: optimizing and repressing rbcM tran-