Differences in LexA regulon structure among Proteobacteria through *in vivo* assisted comparative genomics

Ivan Erill[1ϕ], Mónica Jara[2ϕ], Noelia Salvador[2], Marcos Escribano[1], Susana Campoy[3] and Jordi Barbé[*2,3]

[1] *Biomedical Applications Group, Centro Nacional de Microelectrónica, 08193 Bellaterra, Spain*

[2] *Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona 08193 Bellaterra, Spain*

[3] *Centre de Recerca en Sanitat Animal (CReSA), 08193 Bellaterra, Spain*

ABSTRACT

The LexA regulon encompasses an ensemble of genes involved in preserving cell viability under massive DNA damage and is present in most bacterial phyla. Up to date, however, the scope of this network had only been assessed in the Gamma Proteobacteria. Here we report the structure of the LexA regulon in the Alpha Proteobacteria, using a combined approach that makes use of *in vitro* and *in vivo* techniques to assist and validate the comparative genomics *in silico* methodology. This leads to the first experimentally validated description of the LexA regulon in the Alpha Proteobacteria, and comparison of regulon core structures in both classes suggests that a least common multiple set of genes (*recA*, *ssb*, *uvrA* and *ruvCAB*) might be a defining property of the Proteobacteria LexA network.

INTRODUCTION

Preservation of genetic material is one of the main functions of living beings, and it is perhaps in bacteria where the mechanisms for DNA preservation have been more clearly identified and studied. A global mechanism to respond to DNA lesions (the SOS system) was first described (1) and has been extensively studied (2, 3) in the enteric Gamma Proteobacteria *Escherichia coli*. The SOS response of *E. coli* comprises the DNA damage-mediated induction of at least 40 genes involved in DNA repair and cell survival (2, 3) and is regulated by the LexA and RecA proteins.

[ϕ] Ivan Erill and Monica Jara should be regarded as joint first authors in this work.

[*] To whom correspondence should be addressed at: Departament de Genètica i Microbiologia, Ed. Cn, Universitat Autònoma de Barcelona, 08193 - Bellaterra, Spain. jordi.barbe@uab.es.

Under normal circumstances, *E. coli* LexA represses the expression of SOS genes by specifically binding to a palindromic motif ($CTGTN_8ACAG$) in their promoter: the SOS box (1). In the advent of DNA damage, RecA acquires an active state ($RecA^*$) through binding to single-stranded regions of DNA generated by either DNA damage-mediated replication inhibition or enzymatic processing of broken DNA ends (4). The $RecA^*$ complex then promotes the autocatalytic cleavage of the Ala^{84}-Gly^{85} bond of *E. coli* LexA (5). This cleavage, similar to that carried out by serine proteases (5, 6), renders LexA unable to bind SOS regulatory motifs and, thereby, results in a global induction of the SOS response. Once DNA lesions have been repaired, the intracellular concentration of $RecA^*$ diminishes as new RecA is promptly produced due to SOS induction. Non-cleaved LexA, which is also induced by the SOS response, returns rapidly to normal levels, repressing again the SOS genes and itself.

So far, presence of the *lexA* gene has been reported in almost all bacterial phyla, and distinct LexA-binding motifs have been described for different bacterial phyla and classes. The Gram-positive, for instance, present a highly conserved LexA recognition motif with consensus sequence CGAACRNRYGTTYC (7, 8) that is highly similar to that reported for Cyanobacteria (RGTACNNNDGTWCB; 9). Then again, the LexA recognition sequence of *E. coli* has been reported in several Gamma Proteobacteria families (e.g. *Pseudomonaceae*, *Aeromonadaceae* or *Vibrionaceae*) and in some Beta Proteobacteria (e.g. *Ralstonia solanacearum*; 12), while a markedly different LexA binding motif, a direct repeat with consensus sequence $GAACN_7GAAC$ or $GTTCN_7GTTC$, has been described in the Alpha Proteobacteria class (10, 11) that comprises, among other, the *Caulobacterales*, the *Rhizobiales* and the *Rhodobacterales* orders. Interestingly, all reported LexA binding motifs are monophyletic for the phyla and classes presenting them, suggesting that they may be reliable indicators of branching points in the evolution of bacteria (12).

In contrast to LexA-binding sequences, little is known about the composition of the LexA regulon beyond *E. coli*. *In silico* analyses have shown that a LexA regulated SOS network with the *E. coli* SOS box is present in all the Gamma Proteobacteria sequenced so far and in some Beta Proteobacteria (12). In all these species LexA controls a gene network related to that of *E. coli*, which comprises error prone polymerases (*umuDC*, *dinP*), recombinases (*recA*, *recN*), excision repair nucleases (*uvrAB*) and helicases (*uvrD*) and a cell-division inhibitor (*sulA*).

However, and in spite of the experimentally reported presence of LexA and of some regulated genes in the Gram-positive bacteria (13, 14), the Cyanobacteria (9) and the Alpha Proteobacteria (11), no systematic analyses of the LexA regulon structure in these phyla has been carried out so far. Still, indications on regulon composition are of substantial interest because they can pinpoint subtler differences between species than regulatory motifs (12, 15) and because they can yield hints on how the nature and function of the SOS response may have been shaped across different phyla and in response to particular environments.

In recent years, the increasing availability of sequenced genomes has fostered the design of algorithms to predict regulatory binding sites and thus extend the knowledge on or discover new regulatory networks through *in silico* analyses (2, 17). Based on different statistical approaches, consensus-building (18), expectation maximization (19), oligonucleotide-frequency analysis (20) and Gibbs-sampling method (21) algorithms have been devised to locate new regulatory sites. Simple *in silico* screening, though, is too inaccurate to extract solid knowledge if it is not assisted by prior experimental knowledge on the nature of the regulon (22), thus limiting the application scope of such analyses. More recently, and with the assumption that gene networks and regulatory motifs ought to be dependably conserved across related species (15), comparative genomics analyses have been carried out (23, 24) making use of known regulon structures in related genomes as a means to strengthen and focus motif-prediction algorithms in previously unstudied species. However, even with the comparative genomics approach, extensive experimental knowledge of the regulon under study must still be available in closely related species in order to derive conclusive facts.

In this work, we have made use of a consensus-building algorithm (12, 18) to conduct a comparative genomics analysis of the LexA regulon of Alpha Proteobacteria. Based on prior experimental data (11) and on the known structure of the Gamma Proteobacteria SOS regulon (12), the analysis has been refined through experimental validation of its preliminary results, thus circumventing the lack of extensive experimental knowledge of the LexA regulon in an Alpha Proteobacteria species, to achieve the first consistent outline of the SOS response network in this bacterial class. These results, together with previously published thorough analyses of both the *E. coli* (2) and Gamma Proteobacteria (12) LexA regulons, allow for the first time a direct comparison of the LexA regulon between different Proteobacteria classes. Such a straight comparison is particularly

appealing because, apart from the established phylogenetic divergence between the Alpha and Gamma Proteobacteria (25, 26), both these classes have been shown also to present markedly divergent LexA-binding motifs (11, 12).

MATERIALS AND METHODS

Bacterial strains and growth conditions

The *Sinorhizobium meliloti* 2021 strain used in the present work was grown at 30°C in LB medium (27). All plasmid constructions and cloning experiments were performed in *E. coli* DH5α using the pGEM-T vector. Plasmid DNA was transformed into competent *E. coli* cells as described previously (28).

Nucleic acids techniques

RNA and DNA total extraction was carried out by standard methods (27). Genes and promoter fragments for electrophoretic mobility shift assays were isolated by PCR from total DNA extraction, using suitable oligonucleotide primers designed in accordance to the *S. meliloti* published sequence. RT-PCR (Reverse Transcription PCR) analyses of gene expression were performed for all genes as reported (29), using specific internal oligonucleotide primers for each one. In all cases, the RNA concentration of the gene to be analyzed was always normalized to that of the *S. meliloti trpA* gene, since expression of the latter is not affected by DNA damage (3).

Purification of LexA protein

The *S. meliloti lexA* gene was cloned by PCR using specific primers designed from its published sequence. The 5'-end of the upper primer contained an *Nde*I restriction site in which the ATG initial triplet of the *lexA* gene was included. The lower primer started 200 bp downstream of the translational stop codon of the *lexA* gene. PCR fragment containing the *S. meliloti lexA* gene was cloned into a pGEM-T vector and, afterwards, inserted into a pGEX4T1 expression vector. pGEX4T1-derivative containing the *S. meliloti lexA* gene was transformed into the *E. coli lexA* (Def) BL21(DE3) codon plus strain (2) for over-expression of its encoding LexA protein, which was subsequently purified using the Talon ™ Metal Affinity Resin Kit (Clontech) as reported (9). *S. meliloti* LexA protein obtained was

above 95 % purity as determined with Coomassie Blue staining of SDS-PAGE (15 %) polyacrylamide gels (data not shown) following standard methodology (30).

Electrophoresis mobility shift assays

LexA-DNA binding was analyzed for each gene promoter by electrophoresis mobility shift assays (EMSAs) using purified *S. meliloti* LexA protein. DNA probes were prepared by PCR amplification with one of the primers labeled at its 5' end with digoxigenin (DIG) and purifying each product in a 2% -3% low-melting-point agarose gel. DNA-protein reactions (20 μl) typically containing 20 ng of the DIG-DNA-labeled probe and 80 nM of purified LexA protein were incubated in binding buffer: 10 mM N-2-Hydroxyethyl-piperazine-N' 2- ethanesulphonic acid (HEPES), NaOH (pH 8), 10 mM Tris-HCl (pH 8), 5% glycerol, 50 mM KCl, 1 mM EDTA, 1 mM DTT, 1mg ml-1 of salmon DNA and 50 μg/ml BSA. After 30 minutes at 30ºC, the mixture was loaded onto a 6% non-denaturing Tris-glycine polyacrylamide gel (pre-run for 30 minutes at 10 V/cm in 25mM Tris-HCl (pH 8.5), 250 mM glycine, 1 mM EDTA). DNA-protein complexes were separated at 150 V for 60 min, followed by transfer to a Biodine B nylon membrane (Pall Gelman Laboratory). DIG-labeled DNA-protein complexes were detected following the manufacturer protocol (Roche). For the binding-competition experiments, a 300-fold molar excess of either specific or nonspecific-unlabelled competitor DNA was also included in the mixture. All EMSAs were repeated a minimum of three times to ensure reproducibility of results.

Genome sequences

Available complete genome sequences for the Alpha Proteobacteria species here analyzed were obtained from the from the NCBI Entrez genomes database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome) or from The Institute for Genomic Research (TIGR) Microbial Genome Database (http://www.tigr.org/tdb/mdb/mdb.html).

***In silico* analyses**

In silico analyses of regulon structure were carried out using RCGScanner (Recursive Comparative Genome Scanner), a consensus-building software for the prediction of regulatory motifs that has been previously described (12). Essentially,

the program scans a local raw genome file searching for direct or inverted repeats in the vicinity of putative Open Reading Frames (ORF). After scanning, the program filters out sequences according to their Heterology Index (HI; 31), using both direct cut-off and iterative filtering techniques. NCBI Genbank database is then queried through BLAST (32) to obtain functional definitions for the ORFs that are adjacent to filter-passing motifs. RCGScanner allows two different modes of operation, depending on the availability of experimental information concerning the regulon and organism under study. If such information is available, in the form of known regulatory motifs, RCGScanner uses these motifs to directly generate the consensus sequence that is applied in filtering. Conversely, if no binding motifs are known for the species under study, the program takes as input a known regulon structure in the form of regulon genes sequences. Gene homologues are then searched for through BLAST in the genome of the species under study and, if found, putative regulatory motifs are sought in their promoters. These putative motifs are then used to create the consensus sequence for filtering.

RESULTS AND DISCUSSION

Initial analysis of the Alpha Proteobacteria LexA regulon

Since the structure of the LexA regulon has only been clearly defined in *E. coli* (2) and close relatives (12), experimental validation of *in silico* results was necessary to elucidate the structure of the LexA regulon in the distant Alpha Proteobacteria, which present a markedly divergent LexA-binding motif (11). Therefore, a two-step analytical procedure was implemented as described previously (12). In the initial analysis, the consensus-building software was run against Alpha Proteobacteria complete genomes using the *E. coli* LexA regulon structure as input. Protein sequences of genes that are known to form part of the *E. coli* SOS network (2) were automatically searched for in the analyzed genomes using BLAST and a minimum identity level of 60% as threshold. The promoter regions of the resulting conserved genes were then scanned for putative $GTTCN_7GTTC$ or $GAACN_7GAAC$ direct repeats, and these were used to build a preliminary consensus matrix for filtering.

The results of this initial analysis (Table 1) revealed that a LexA core regulon structure (*lexA*, *recA*, *uvrA* and *ssb*) similar to that of Gamma Proteobacteria (12) might be present in the Alpha Proteobacteria. Using the aforementioned consensus

matrix, motifs were filtered using astringent selection criteria (12). For each motif putatively regulating a given gene, these criteria impose a HI score below 6 and the presence of a motif upstream of a homologue of that same gene in at least another bacterial species. After filtering, several high-scoring LexA-biding sites upstream of contrasted SOS gene homologues (*lexA*, *recA*, *uvrA*, *ssb*, *sulA* and *dinP*) were identified in almost all the Alpha Proteobacteria genomes analyzed, as well as upstream of some DNA-repair associated genes (*ruvC* and *dnaE*) that had not been previously described as LexA-regulated in either Alpha or Gamma Proteobacteria. The only exceptions to this trend were the intracellular parasites *Rickettsia conorii* and *Rickettsia prowazekii*, which present a deletion of their *lexA* gene due to drastic genome reduction. The thus selected LexA-binding motifs, together with experimentally determined LexA boxes of several Alpha Proteobacteria (11, Table 1), were then used to define a robust interspecies consensus sequence (Figure 1) to carry out a second, more accurate filtering step.

Experimental validation of the *in silico* approach

Prior to conducting a full-fledged analysis of the Alpha Proteobacteria LexA regulon using the interspecies consensus sequence obtained in the initial analysis, a pilot study was carried out in the nodule-forming soil bacterium *S. meliloti* to validate the reliability of the *in silico* approach. Of the 29 *S. meliloti* genes presenting at least one putative regulatory motif with HI<6 in this second round of filtering (Table 2), 6 of those not previously reported to be DNA damage-inducible in the Alpha Proteobacteria (*ruvC*, *dinP*, *sulA1*, *parE*, *yigN* and SMc03093), together with *lexA*, were arbitrarily elicited for experimental validation.

After cloning and purifying the LexA protein of *S. meliloti*, EMSAs were carried out to determine the LexA-binding affinity of promoters for the 6 chosen genes. Results (Figure 2a) clearly demonstrated that all 6 promoters are able to bind LexA, suggesting that they might be DNA-damage inducible genes. To further elucidate this point, RNA was extracted from *S. meliloti* cultures following exposure to mitomycin C and analyzed through RT-PCR. Again, the results (Figure 2b) clearly established that all 6 genes were DNA-damage inducible, confirming that the two-step *in silico* approach here taken and the use of a robust interspecies consensus yielded manifestly reliable results.

Analysis of the Alpha Proteobacteria LexA regulon

After both *in vitro* and *in vivo* validation of the filtering scheme taken in *S. meliloti*, the second round of analyses was extended to the remaining Alpha Proteobacteria species with published complete genomes (Table 3). To avoid false positives, a combined astringent filtering procedure was applied, including only those genes that, apart from presenting at least one motif with a HI<6, were seemingly regulated in at least three different bacterial species.

The results (Table 3) allowed extending the preliminary definition of the Alpha Proteobacteria LexA regulon core (*lexA*, *recA*, *ssb* and *uvrA*) by imposing the more severe criterion of presence in at least 5 of the 10 bacterial species analyzed. The thus identified core encompasses 13 genes that include previously described Alpha and Gamma Proteobacteria LexA-regulated genes (*lexA*, *recA*, *ssb* and *uvrA*), several *E. coli* SOS genes (*dinP*, *yigN* and *sulA*) and some new LexA-regulated genes identified here (*parE*, *dnaE*, *ruvC*, *ispE*, SMc00865 and *comM*).

Again, to confirm the validity of these *in silico* results, those members of the identified regulon core that had not been previously experimentally validated (*dnaE*, *sulA2*, *ispE*, SMc00865 and *comM*) were analyzed in *S. meliloti*. EMSA results (Figure 3a) demonstrate that all these genes are able to bind the LexA protein in *S. meliloti*. Furthermore, subsequent RT-PCR analyses (Figure 3b) revealed that these genes are DNA-damage inducible in *S. meliloti*, demonstrating that the *in silico* identified regulon core is indeed functional in this bacterial species. In addition, three other genes (*ppdK*, *dnrV* and *recG*) presented high-scoring LexA-binding motifs in at least three different bacterial species, and were thus considered as optional members of the LexA regulon in the Alpha Proteobacteria. As expected, and in agreement with the results of the initial analysis, there was again no evidence of LexA regulatory motifs in the *Rickettsiae*, indicating that the loss of *lexA* must have taken place early in the evolution of these intracellular parasites, and that subsequent genome reduction has removed all traces of former LexA regulation.

It should be stressed that these results constitute also the first description of a *sulA*-like gene under control of the LexA protein in the Alpha Proteobacteria class. Moreover, the surrounding region of the two copies (three in the case of *A. tumefaciens*) identified here of this LexA-regulated *sulA* homologue presents the same genetic organization in all the species analyzed. This genetic arrangement, consisting of the own *sulA* homologue, a DNA polymerase IV homologue (*dinP*) and a homologue of the alpha subunit of DNA polymerase III (*dnaE*), has been shown

to be a polycistronic transcriptional unit belonging to a broader class of mobile genetic element encoding also the LexA protein in a *lexA-sulA-dinP-dnaE* cassette organization (33) that is present in some Gamma Proteobacteria. The presence of a cell-division inhibitor homologue in the Alpha Proteobacteria LexA regulon supports the view that postponing cell division under massive DNA damage is a markedly favorable and widespread adaptation, a hypothesis further endorsed by the reported convergent evolution of a similar mechanism mediated by the *yneA* gene in the Gram-positive bacterium *Bacillus subtilis* (34).

Comparison of the Alpha and Gamma LexA regulons

The previously described regulon core for the Gamma Proteobacteria (12) consists of 6 genes besides *lexA* whose regulation seems conserved across almost all the species analyzed to date. These genes encode, respectively, the DNA strand exchange and recombination protein RecA, both Holliday junction helicase subunits A and B (RuvAB), the single-strand binding protein Ssb, the recombination protein RecN and the excision nuclease subunit A (UvrA). In the case of the Alpha Proteobacteria regulon core, a highly similar structure is present, with *recA*, *ssb* and *uvrA* explicitly regulated, and the *ruvAB* regulation substituted in this case by the regulation of the equivalent *ruvCAB* operon present in all the Alpha Proteobacteria species here analyzed but the Rickettsiae. Taking this substitution into account, the only protein of the Gamma Proteobacteria regulon core that is missing in the Alpha Proteobacteria one is the recombination protein RecN. In light of the significant phylogenetic divergence between Alpha and Gamma Proteobacteria, such a high degree of similarity in regulon core composition suggests the possibility that there is a least common multiple set of genes that make up the LexA regulon of Proteobacteria: *recA*, *uvrA*, *ssb* and the *ruvAB*/*ruvCAB* operon. The definition of such a least common multiple is interesting because it can contribute to reveal the common evolutionary pressures that maintain the essence of the LexA regulon in different bacterial classes. This line of reasoning is further strengthened when different reports confirming LexA regulation of some of these same genes in the Gram-positive Phylum [*recA* (7), *uvrA* (35), *ssb* (36) or *ruvC* (14)] are taken into account. Thus, similar studies in the near future could reveal a universal least common multiple set of genes for the bacteria LexA regulon, shedding more light on the general mechanisms governing the evolution of the LexA and other complex regulons.

Another interesting point of the straight comparison between regulon cores concerns the additions to the Alpha Proteobacteria regulon core. These additions are significant because they can pinpoint shared evolutionary pressures and are indicative of the flexibility of the LexA regulon in co-opting additional genes. Of the 8 additions to the Alpha Proteobacteria LexA regulon core with respect to its Gamma Proteobacteria counterpart, some can be readily explained by their reported involvement in DNA repair or in the overall SOS response of *E. coli* and other Gamma Proteobacteria. This is the case for the aforementioned *sulA* homologue (16), the DNA polymerase IV *dinP* (37), the alpha subunit of DNA polymerase III (*dnaE*; 38) and the hypothetical protein *yigN* (39). The presence of the DNA topoisomerase IV subunit B (*parE*) could be similarly explained by its reported involvement in mutagenic processes and antimicrobial resistance (40, 41). Regarding the other three additions, however, it is difficult to derive sound inferences without further experimental work on their respective protein functions. The *comM* gene, for instance, has been annotated as a Mg^{2+} chelatase in *B. suis* and *B. melitensis*, and as such it seems feasible that it could be involved in the regulation of polymerase fidelity during the SOS response through the sequestering of magnesium (42). In a different setting, the *ispE* gene here reported has been annotated as the molybdenum cofactor biosynthesis protein A (*moaA*) in *C. crescentus*, a gene that has been linked in *E. coli* to the detoxifying processes ensuing N-6-hydroxylaminopurine (HAP) induced lesions (43). Therefore, it seems not farfetched to assume that some environmental factor may have fostered its co-option in the LexA regulon of the Alpha Proteobacteria.

Conclusion

In the present work, we have made use of experimental validation to make a robust assessment of regulon structure for a whole bacterial class through comparative genomics. The inclusion of an intermediate experimental stage improves the accuracy of the consensus-building method used and adds a layer of reliability to the results obtained through comparative genomic approaches. This allows extending the range of comparative genomics assays to different bacterial classes with markedly divergent regulatory motifs, as in the case of Gamma and Alpha Proteobacteria LexA regulons. Using this approach, we have analyzed the LexA regulon of Alpha Proteobacteria, providing the first comprehensive description of the LexA regulon in this bacterial class. The results show that a

least common multiple set of genes may be the norm in the Proteobacteria LexA regulon, and reveal some interesting additions to the LexA regulon of Alpha Proteobacteria that may be linked to their particular environment and evolution.

Acknowledgements

This work was funded by Grants BMC2001-2065 and BFM2004-02768/BMC from the Ministerio de Educación y Ciencia (MEC) de España and 2001SGR-206 from the Departament d'Universitats, Recerca i Societat de la Informació (DURSI) de la Generalitat de Catalunya, and by the Consejo de Investigaciones Científicas (CSIC). M. Jara was recipient of a pre-doctoral fellowship from the DURSI and S. Campoy is recipient of a post-doctoral contract from INIA-IRTA. We are deeply indebted to Joan Ruiz and Dr. Pilar Cortés for their excellent technical assistance.

References

1. Walker,G.C. (1984) Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. *Microbiol Rev.* **48**, 60–93.
2. Fernández De Henestrosa,A.R., 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.
3. Courcelle,J., Khodursky,A., Peter,B., Brown,P.O., Hanawalt,P.C. (2001) Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics* **158**, 41-64.
4. Sassanfar,M., and Roberts,J.W. (1990) Nature of SOS-inducing signal in *Escherichia coli*. The involvement of DNA replication. *J Mol Biol* **212**, 79–96.
5. Little,J.W. (1991) Mechanism of specific LexA cleavage: autodigestion and the role of RecA coprotease. *Biochimie* **73**, 411–421.
6. Luo,Y., Pfuetzner,R.A., Mosimann,S., Paetzel,M., Frey,E.A., Cherney,M., Kim,B., Little,J.W., and Strynadka,N.C. (2001) Crystal structure of LexA: a conformational switch for regulation of self-cleavage. *Cell* **106**, 585–594.
7. Winterling,K.W., Chafin,D., Hayes,J.J., Sun,J., Levine,A.S., Yasbin,R.E. and Woodgate,R. (1998) The *Bacillus subtilis* DinR binding site: redefinition of the consensus sequence. *J. Bacteriol.* 180, 2201-2211.
8. Davis,E.O., Dullaghan,E.M. and Rand,L. (2002) Definition of the Mycobacterial SOS box and use to identify LexA-regulated genes in *Mycobacterium tuberculosis*. *J. Bacteriol*. **184**, 3287–3295.

9. Mazón,G., Lucena,J.M., Campoy,S., Fernández de Henestrosa,A.R., Candau,P. and Barbé,J. (2003) LexA-binding sequences in Gram-positive and cyanobacteria are closely related. *Mol. Gen. Genomics* **271**, 40-49.
10. Fernández de Henestrosa,A.R., Rivera,E., Tapias,A. and Barbé J. (1998) Identification of the *Rhodobacter sphaeroides* SOS box. *Mol. Microbiol.* **28**, 991-1003.
11. Tapias,A. and Barbé,J. (1999) Regulation of divergent transcription from the *uvrA-ssb* promoters in *Sinorhizobium meliloti*. *Mol. Gen. Genet.* **262**, 121–130.
12. Erill,I., Escribano,M., Campoy,S., and Barbé,J. (2003) In silico analysis reveals substantial variability in the gene contents of the Gamma Proteobacteria LexA-regulon. *Bioinformatics* **19**, 2225-2236.
13. Mostertz,J., Scharf,C., Hecker,M., Homuth,G. (2004) Transcriptome and proteome analysis of *Bacillus subtilis* gene expression in response to superoxide and peroxide stress. *Microbiology* **150**, 497-512.
14. Davis,E.O., Dullaghan,E.M., Rand,L. (2002) Definition of the mycobacterial SOS box and use to identify LexA-regulated genes in *Mycobacterium tuberculosis*. *J Bacteriol.* **184**, 3287-3295.
15. Gelfand,M.S., Novichkov,P.S., Novichkova,E.S., Mironov,A.A. (2000) Comparative analysis of regulatory patterns in bacterial genomes. *Brief. Bioinform.* **1**, 357-371.
16. Hill,T.M., Sharma,B., Valjavec-Gratian,M. and Smith,J. (1997) *sfi*-independent filamentation in *Escherichia coli* is *lexA* dependent and requires DNA damage for induction. *J. Bacteriol.* **179**, 1931-1939.
17. Rodionov,D.A., Mironov,A.A. and Gelfand,M.S. (2001) Transcriptional regulation of pentose utilisation systems in the *Bacillus/Clostridium* group of bacteria. *FEMS Microbiol. Lett.* **205**, 305-314.
18. Stormo,G.D. and Hartzell,G.W. (1989) Identifying protein-binding sites from unaligned DNA fragments. *Proc. Natl. Acad. Sci. USA*, **86**, 1183-1187.
19. Lawrence,C.E. and Reilly,A.A. (1990) An EM Algorithm for the Identification and Characterization of Common Sites in Unaligned Biopolymers Sequence. *Proteins*, **7**, 41-51.
20. van Helden,J., André,B. and 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.
21. Lawrence,C.E., Altschul,S.F., Boguski,M.S., Liu,J.S., Neuwald,A.F. and Wooton,J.C. (1993) Detecting subtle sequence signals: A Gibbs sampling strategy for multiple alignment. *Science*, **262**, 208-214.
22. Bailey,T.L. and Elkan,C. (1995) The value of prior knowledge in discovering motifs with MEME, *Proc. IIIrd Int. Conf. Intel. Sys. Mol. Biol*, 21-29, AAAI Press, Menlo Park, California.

23. Panina,E.M., Vitreschak,A.G., Mironov,A.A. and Gelfand, M.S. (2003) Regulation of biosynthesis and transport of aromatic amino acids in low-GC Gram-positive bacteria. *FEMS Microbiol Lett.* **222**, 211-20.
24. Yellaboina,S., Seshadri,J., Kumar,M.S. and Ranjan,A. (2004) PredictRegulon: a web server for the prediction of the regulatory protein binding sites and operons in prokaryote genomes. *Nucleic Acids Res.* **32**, W318-320.
25. Eisen,J.A. (1995) The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J. Mol. Evol.*, **41**, 1105-1123.
26. Woese,C.R. and Fox,G.E. (1977) Phylogenetic Structure of the Prokaryotic Domains: The Primary Kingdoms. *Proc. Natl. Acad. Sci. USA*, **74**, 5088-5090.
27. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning. A Laboratory Manual*, 3rd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
28. Silhavy,T.J., Berman,M.L. and Enquist,L.W. (1984) *Experiments with Gene Fusion*. Cold Spring Harbor Laboratory. NY: Cold Spring Harbor.
29. Campoy,S., Fontes,M., Padmanabhan,S., Cortes,P., Llagostera,M. and Barbé,J. (2003) LexA-independent DNA damage-mediated induction of gene expression in *Myxococcus xanthus*. *Mol. Microbiol.* **49**, 769-781.
30. Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
31. Berg,O.G. and von Hippel,P.H. (1987) Selection of DNA binding sites by regulatory proteins, statistical-mechanical theory and application to operators and promoters. *J. Mol. Biol.* **193**, 723-750.
32. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.
33. Abella,M., Erill,I., Jara,M., Mazón,G., Campoy,S. and Barbé,J. (2004) Widespread distribution of a *lexA*-regulated DNA damage-inducible multiple gene cassette in the Proteobacteria phylum. *Mol. Microbiol.* **54**, 212-222.
34. Kawai,Y., Moriya,S. and Ogasawara,N. (2003) Identification of a protein, YneA, responsible for cell division suppression during the SOS response in *Bacillus subtilis*. *Mol. Microbiol.* **47**, 1113-1122.
35. Smith,B.T., Grossman,A.D. and Walker,G.C. (2002) Localization of UvrA and effect of DNA damage on the chromosome of *Bacillus subtilis*. *J. Bacteriol.* **184**, 488-493.
36. Lindner,C., Nijland,R., van Hartskamp,M., Bron,S., Hamoen,L.W. and Kuipers,O.P. (2004) Differential expression of two paralogous genes of *Bacillus subtilis* encoding single-stranded DNA binding protein. *J. Bacteriol.* **186**, 1097-1105.

37. Wagner,J., Gruz,P., Kim,S.R., Yamada,M., Matsui,K., Fuchs,R.P. and Nohmi,T. (1999) The *dinB* gene encodes a novel *E. coli* DNA polymerase, DNA pol IV, involved in mutagenesis. *Mol Cell.* **4**, 281-286.
38. Timms,A.R. and Bridges,B.A. (2002) DNA polymerase V-dependent mutator activity in an SOS-induced *Escherichia coli* strain with a temperature-sensitive DNA polymerase III. *Mutat. Res.* **499**, 97-101.
39. Van Dyk,T.K., DeRose,E.J. and Gonye,G.E. (2001) LuxArray, a high-density, genomewide transcription analysis of *Escherichia coli* using bioluminescent reporter strains. *J. Bacteriol.* **183**, 5496-5505.
40. Ferrero,L., Cameron,B. and Crouzet,J. (1995) Analysis of *gyrA* and *grlA* mutations in stepwise-selected ciprofloxacin-resistant mutants of *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 39, 1554-1558.
41. Oh,H., Stenhoff,J., Jalal,S. and Wretlind,B. (2003) Role of efflux pumps and mutations in genes for topoisomerases II and IV in fluoroquinolone-resistant *Pseudomonas aeruginosa* strains. *Microb. Drug Resist.* **9**, 323-328.
42. Yang,L., Arora,K., Beard,W.A., Wilson,S.H. and Schlick,T. (2004) Critical role of magnesium ions in DNA polymerase beta's closing and active site assembly. *J. Am. Chem. Soc.* **126**, 8441-8453.
43. Burgis,N.E., Brucker,J.J. and Cunningham,R.P. (2003) Repair system for noncanonical purines in *Escherichia coli*. *J. Bacteriol.* **185**, 3101-3110.
44. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: A sequence logo generator, *Genome Res.* **14**, 1188-1190.

TABLES

Gene	Motif	HI	
AT_*lexA*	GAACACATATGGAAC	0,70	
AT_*recA*	AAACGAAAGCAGAAC	7,44	*
AT_*recA*	GAACAAATAGAGTAC	2,07	*
AT_*ssb*	GAACAAAAAAGGAAC	1,50	
AT_*uvrD*	GAATAAAAGCAGAAC	3,40	
AT_*ruvC*	GAACAAAACGACAAC	4,95	
AT_*dnaE*	GAACAAAATGAGAAC	2,00	
AT_*dnaE*	GAACAAAGTTGGAAC	4,32	
BM_*lexA*	GAACAAGACTGGAAC	1,67	
BM_*lexA*	AAACCATTGCAGAAC	2,18	
BM_*recA*	GAACAAGAATGGAAC	2,77	
BM_*ssb*	GAACAAAACAGGAAC	1,49	
BM_*uvrD*	GCACACCGGCTGAAC	3,17	
BS_*lexA*	GAACAAGACTGGAAC	1,83	
BS_*lexA*	AAACCATTGCAGAAC	3,61	
BS_*recA*	GAACAAGAATGGAAC	2,42	
BS_*ssb*	GAACAAAACAGGAAC	1,58	
BS_*sulA*	GAACATAAAGTGAAC	3,60	
CC_*lexA*	GAACACCAGGAGAAC	0,57	
CC_*recA*	GAACAAAGAGTGTAC	5,17	
CC_*recA*	GAACATCTTGCGAAC	4,42	
CC_*ssb*	GAACGTTATGAGAAC	3,12	
CC_*uvrA*	GAACGTCGCGAGAAC	2,33	
CC_*uvrD*	AAACGCTCGGTGAAC	3,98	
ML_*lexA*	AAACAGTTGCAGAAC	0,34	
ML_*recA*	GAACAAAAAAGGTAC	3,29	
ML_*recA*	GTACGAAAAAAGAAC	5,10	
ML_*ssb*	GAACGAAAAGGGAAC	3,04	
ML_*sulA*	GAACATAACAGGAAC	4,31	
SM_*lexA*	GAACACATATGGAAC	1,18	*
SM_*recA*	GAACAAGAATCGAAC	0,40	*
SM_*recA*	GAACAAAACATGTAC	3,56	*
SM_*ssb*	GAACAAAAAAAGAAC	1,04	*
SM_*uvrD*	GAATAAAAGAAGAAC	3,05	
SM_*dnaE*	GAACAAAAAGGGAAC	1,38	
SM_*dnaE*	GAACACGCAGTAAAC	3,22	
SM_*dnaE*	GAACGGAAATAGAAC	4,08	
SM_*sulA*	GAACATAACATGAAC	2,93	
RS_*recA*	GAACATAGGGCGAAC	ND	*
RC_*recA*	GAACAAGACAGGAAC	ND	*
RE_*recA*	AAACAAATATAGAAC	ND	*
RE_*recA*	GAACAAATAGGGTAC	ND	*
RV_*recA*	GAACAAATCGTGTAC	ND	*
PD_*recA*	GAACAACCCGTGAAC	ND	*
AF_*recA*	GTACGTTGACAAAAC	ND	*
BA_*ssb*	GAACAAAACAGGAAC	ND	*
BA_*recA*	GAACAAGAATGGAAC	ND	*

Gene	Motif	HI	
AT_*lexA*	GTTCTGTATTTGTTT	0,90	
AT_*uvrA*	GTTCCTTTTTTGTTC	1,55	
AT_*uvrD*	GTTCAGCATTTGTTC	3,05	
AT_*dinP*	GTTCTGGTTTTGTTT	1,24	
AT_*ruvC*	GTTGTCGTTTTGTTC	4,72	
BM_*lexA*	GTTCTGGTTTTGTTT	1,38	
BM_*recA*	GTTCGTGGATAGTTC	4,23	
BM_*uvrA*	GTTCGATATTTGTTC	3,58	
BM_*dinP*	GTTCCTTTTATGTTC	3,75	
BM_*ruvC*	GTTTCTCTTTTGTTC	1,46	
BS_*lexA*	GTTCTGGTTTTGTTT	1,54	
BS_*recA*	GTTCGTGGATAGTTC	3,21	
BS_*uvrA*	GTTCGATATTTGTTC	3,83	
BS_*ruvC*	GTTTCTCTTTTGTTC	1,54	
CC_*lexA*	GTTTGCGGTTTGTTC	1,44	
CC_*uvrA*	GTTCGCATCTTGTTC	2,20	
CC_*uvrD*	GTGCTACATATGTTC	4,92	
ML_*lexA*	GTTCTGGGTTTGTTT	1,22	
ML_*uvrA*	GTTCGGCCTTTGTTC	2,97	
ML_*ruvC*	GTTTCCGGTTTGTTC	3,00	
SM_*lexA*	GTTCTTGATTTGTTT	1,44	
SM_*uvrA*	GTTCTTTTTTTGTTC	1,07	*
SM_*dinP*	GTTCAACATTTGTTC	3,18	
SM_*ruvC*	GTTTTTGTTTTGTTC	1,89	
RS_*recA*	GTTCGCCTTATGATC	ND	*
RS_*uvrA*	GTTCATACTATGTTC	ND	*
RC_*recA*	GTTCCGAAATTGTTC	ND	*
RC_*uvrA*	GTTCCTGTTCCGTTC	ND	*
RV_*recA*	GTTCTCTTCTTGTTC	ND	*
BA_*recA*	GTTCGTGGATAGTTC	ND	*
BA_*uvrA*	GTTCGATATTTGTTC	ND	*
AM_*recA*	GTTCTCCTCTCGTTC	ND	*
PD_*uvrA*	GTTCCTGTGATGTTC	ND	*

Table 1. Regulatory motifs used to build the interspecies consensus for the second phase of the analysis. The Heterology Index (HI) score is displayed for all the motifs detected in the initial analysis of available complete genomes, while *ND* indicates non-determined HI scores. An asterisk (*) denotes those motifs that have been experimentally described to be involved in LexA regulation in the Alpha Proteobacteria (11). Motifs in homologues of known *E. coli* SOS genes (2) are shaded in grey. Name abbreviations are as follows: AF – *Acidiphilium facilis*, AM – *Aquaspirillum magnetotacticum*, AT – *Agrobacterium tumefaciens*, BA – *Brucella abortus*, BM – *Brucella melitensis*, BS – *Brucella suis*, CC – *Caulobacter crescentus*, ML – *M. loti*, PD – *Paracoccus denitrificans*, RS – *Rhodobacter sphaeroides*, RC – *Rhodobacter capsulatus*, RE – *Rhizobium etli*, RV – *Rhodopseudomonas viridis*, SM – *S. meliloti*.

Synonym	Name	Position	Strand	Size (bp)	Regulatory sequence	Distance	HI
lexA	SMc01183	1749452	+	714	GTTCTTGATTTGTTT	21	2.297
					GAACACATATGGAAC	47	3.113
recA	SMc00760	1949263	-	1083	GAACAAAACATGTAC	59	3.176
					GAACAAGAATCGAAC	127	2.426
ssb	SMc01233	1689793	+	522	GAACAAAAAAAGAAC	144	0.515
uvrA	SMc01235	1689506	-	2919	GTTCTTTTTTTGTTC	128	0.389
uvrD	SMc01461	2307039	+	2382	GAATAAAAGAAGAAC	234	3.830
ftsK	SMb20596	1610593	-	987	AAACAGAAATTGAAC	182	5.250
ruvC	SMc03967	2961292	-	510	GTTTTTGTTTTGTTC	44	1.745
SMc03968	SMc03968	2961461	+	987	GAACAAAACAAAAAC	111	3.419
dinP	SMc01373	1409085	-	1290	*GTTCCGGATATGATC*	+6	6.547
					GTTCAACATTTGTTC	5	3.666
dnaE	SMc01375	1402647	+	3507	GAACACGCAGTAAAC	102	7.828
					GAACAGTAGCGGAAA	248	9.664
					GAACAAAAAGGGAAC	259	0.001
					GAACGGAAATAGAAC	270	4.147
sulA1	SMc03790	3458293	-	978	GAACATAACATGAAC	+6	2.648
SMc03791	SMc03791	3458436	+	432	GTTCATGTTATGTTC	135	3.134
sulA2	SMa0888	495115	+	897	GAACAAATACAGAAC	+33	2.085
SMa0883	SMa0883	493138	+	528	GTTCCTGCTATGTTC	+26	4.090
SMa0882	SMa0882	493113	-	588	GAACATAGCAGGAAC	38	3.534
ibpA	SMc04040	3044397	-	459	GTTCATCTATTGTTC	156	3.721
					GAACGGCGGCCGAAC	176	9.373
parE	SMc01018	1544311	-	2058	ATTCGCCTTTTGTTC	118	5.592
ispE	SMc00866	921548	-	1155	GTTCTTGATTTGTTC	65	0.719
SMc00865	SMc00865	921727	+	549	GAACAAATCAAGAAC	101	1.929
comM	SMc00420	368205	+	1530	GTTCTATCATTGTTC	139	5.290
SMc00924	SMc00924	859388	-	738	GTTTCTCTTTTGTTC	255	2.558
yigN	SMc01102	455399	-	1203	GTTCTCGTTTTGATC	35	3.681
SMa0414	SMa0414	224805	+	1668	GTTCCCCCTTTGTTT	285	4.934
					AAACAAATAGGGAAC	152	3.268
SMb20912	SMb20912	1320904	-	1092	GTTCCTATTATGTTC	55	3.524
plsB	SMc02687	2537300	-	828	GTTCGTTTCATGTTA	35	8.378
					GTTCGCTTTTTGTTC	46	1.202
SMc02819	SMc02819	2376785	-	765	GTTCCTGTTTTGTTT	22	1.901
SMc02818	SMc02818	2376977	+	942	AAACAAAACAGGAAC	156	2.711
SMc03093	SMc03093	3256262	-	477	GTTCTTGATTTGTTC	148	0.719

Table 2. *S. meliloti* genes with at least one regulatory motif displaying a HI<6 in the second phase of the analysis. Synonyms are provided for known genes and substituted for annotation loci names elsewhere. The Distance shown is relative to the ORF start codon; a + symbol preceding the distance designates intragenic regulatory sequences. Shaded rows indicate those genes chosen for experimental validation of the *in silico* approach.

	A. tumefaciens				B. japonicum				B. melitensis				B. suis			
Gene	**Name**	**Motif**	**d**	**HI**	**Name**	**Motif**	**d**	**HI**	**Name**	**Motif**	**d**	**HI**	**Name**	**Motif**	**d**	**HI**
lexA	AGR_C_2577	GTTCTGTATTTGTTT	38	3,29	blr4826	GTTCATGATTTGTTT	75	3,72	BMEI0840	GTTCTGGTTTTGTTT	31	2,18	BR1144	GTTCTGGTTTTGTTT	31	2,18
		GAACACATATGGAAC	64	3,11		GAACACATATCGAAC	101	4,12		GAACAAGACTGGAAC	57	1,62		GAACAAGACTGGAAC	57	1,62
						GTGCGCGCGCCGTTC	231	13,09		AAACCATTGCAGAAC	68	9,09		AAACCATTGCAGAAC	68	9,09
recA	AGR_C_3441	AAACGAAAGCAGAAC	+32	5,07	bll5755	GAACAAATAGGGTAC	76	3,06	BMEI0787	GTTCGTGGATAGTTC	+21	5,22	BR1202	GAACAAGAATGGAAC	123	1,42
		GAACAAATAGAGTAC	+100	3,11		GAACATATTGCGAAC	94	5,15		GAACAAGAATGGAAC	+64	1,42		GTTCGTGGATAGTTC	166	5,22
		GTTCGGCAAGGGATC	+263	13,71												
ssb	AGR_C_2789	GAACAAAAAAGGAAC	95	0,47	bll4698	GAACAAATCTGGAAC	154	1,88	BMEI0880	GAACAAAACAGGAAC	188	0,66	BR1102	GAACAAAACAGGAAC	188	0,66
uvrA	AGR_C_2790	GTTCCTTTTTTGTTC	108	0,71	blr4702	GTTCTTCCTACGTTC	99	6,39	BMEI0878	GTTCGATATTTGTTC	142	2,47	BR1104	GTTCGATATTTGTTC	142	2,47
ruvC	AGR_L_2221	GTTGTCGTTTTGTTC	44	3,53	blr1535	GTTCTGTTTCTGTTT	33	5,07	BMEI0332	GTTTCTCTTTTGTTC	46	2,56	BR1704	GTTTCTCTTTTGTTC	46	2,56
dnaE	AGR_C_2379	GAACAAAGTTGGAAC	338	3,63	bll4866	GTTCTTCATACGTTC	351	5,05	BMEI1137	GGACATCGCCCGAAC	158	11,34	BR0825	GGACATCGCTCGAAC	160	10,99
		GAACAAAATGAGAAC	349	1,66		GAACAAAATGAGAAC	369	1,66								
dinP	AGR_C_2382	GTTCTGGTTTTGTTT	+22	2,18	bll0861	GTTCGGCGTTCGCTC	+200	8,79	BMEII0656	GTTCCTTTTATGTTC	20	2,43	BRA0615	GTTCCTTTTATGTTC	22	2,43
	AGR_pAT_692	GTTCTCTCTTTGTTT	49	4,51												
sulA1	AGR_L_3170	GAACAAAACAAGAAC	39	0,71	blr3024	GAACATATCATGAAC	67	3,87	BMEI1874	TAACAGGAATCGAAC	34	9,11	BR0071	TAACAGGAATCGAAC	0	9,11
		GATCATCGGCGGTTC	209	13,43						GAACATAAAGTGAAC	52	1,99		GAACATAAAGTGAAC	18	1,99
sulA2	AGR_pAT_143	GAACAAACAATGAAC	62	3,20												
sulA3	AGR_pTi_172	GAACAAAAACAGAAC	161	0,87												
		GAACAATACTCGTAC	209	4,98												
parE	AGR_C_2992	-			bll4355	-			BMEII0676	GAACGTAAAGCGAAC	185	3,89	BRA0591	GAACGTAAAGCGAAC	129	3,89
ispE	AGR_C_1309	GTTCTTGTATTGTTC	93	1,81	blr2519	GTTCTTGATTTGTTC	4	0,72	BMEI1541	GTTCTTGATTTGTTC	71	0,72	BR0387	GTTCTTGATTTGTTC	71	0,72
						GTTCCTTTGGCGTGC	+7	12,54								
SMc00865	AGR_C_1311	GAACAATACAAGAAC	80	2,18					BMEI1540	GAACAAATCAAGAAC	51	1,93	BR0388	GAACAAATCAAGAAC	51	1,93
comM	AGR_C_541	GTTCTCTAGGTGTTC	79	8,11	bll0661	GTTCCCGTATTGTTC	59	2,62	BMEI1994	-			BR2132	GTTCCTGAAACGTTC	93	6,70
		GTTCTTGTTTTATTC	68	6,06												
		GCTCCCGGTTTGTTC	141	4,32												
yigN	AGR_C_639	GTTCTTGTTTTGATC	+6	3,19	bll8110	GTTCGCTATTCGTTC	53	4,05	BMEII0263	GTTCCTGTTTTGATC	236	3,51	BRA1036	GTTCCTGTTTTGATC	36	3,51
ppdK	AGR_C_1470	-			blr2538	-			BMEI1436	GTTCCCCATTTGTTC	105	2,02	BR0500	GTTCCCCATTTGTTC	235	2,02
					bll2515	GAACATGCCGCGGAC	177	10,09								
dnrV	AGR_C_2825	GTTCGCGTGAATTTC	228	11,37	blr0126	-			BMEI1437	GAACAAATGGGGAAC	+7	1,81	BR0499	GAACAAATGGGGAAC	15	1,81
recG	AGR_C_3275	GAACAGGAGGCGAAC	83	4,61	blr4603	-			BMEII0686	GAACAAAAAGAGCAC	44	2,99	BRA0581	GAACAAAAAGAGCAC	46	2,99
		GACCTCCTATAGAAC	94	14,16												

	C. crescentus				M. loti				*S. meliloti*				R. palustris			
Gene	**Name**	**Motif**	**d**	**HI**	**Name**	**Motif**	**d**	**HI**	**Name**	**Motif**	**d**	**HI**	**Name**	**Motif**	**d**	**HI**
lexA	CC1902	GAACACCAGGAGAAC	12	3,84	mlr0626	GTTCTGGGTTTGTTT	47	3,13	SMc01183	GTTCTTGATTTGTTT	21	2,30	RPA2903	GTTCATGATTTGTTT	+19	3,72
		GTTTGCGGTTTGTTC	53	3,51		GAACACAACTGGAAC	73	2,09		GAACACATATGGAAC	47	3,11		GAACACACCCCGAAC	6	5,51
						AAACAGTTGCAGAAC	84	8,25								
recA	CC1087	GAACAAAGAGTGTAC	90	4,07	mlr0030	GAACAAAAAAGGTAC	59	2,31	SMc00760	GAACAAAACATGTAC	59	3,18	RPA3851	GAACAAATGGGGTAC	69	3,65
		GAACATCTTGCGAAC	108	6,93		GTACGAAAAAAGAAC	125	5,15		GAACAAGAATCGAAC	127	2,43		GAACATATTGCGAAC	87	5,15
		GTTCGGCAAGGGCTC	+108	14,81						*GTTCGGCAAGGGATC*	+104	13,71				
ssb	CC1468	GAACGTTATGAGAAC	87	6,01	mll0743	GAACGAAAAGGGAAC	102	1,57	SMc01233	GAACAAAAAAAGAAC	144	0,52	RPA2814	GAACAAAAATAGAAC	215	0,52
		GTGCTGCAGAAGTTC	+377	11,24						*GAACAATATGTGAAG*	+217	8,47				
uvrA	CC2590	GTTCGCATCTTGTTC	79	4,11	mlr0750	GTTCGGCCTTTGTTC	96	3,47	SMc01235	GTTCTTTTTTTGTTC	128	0,39	RPA2816	GTTCTTCCTATGTTC	83	4,26
		GAACGTCGCGAGAAC	262	6,46		*GAACACAATCTGAAG*	+109	9,25								
ruvC	CC3238	*GAACACATGATGAAC*	+137	4,37	mll3901	GTTTCCGGTTTGTTC	44	3,51	SMc03967	GTTTTTGTTTTGTTC	44	1,75	RPA1099	GTTCTTAAGCTTTTC	19	11,42
														GTTGCTATTTTGTTC	30	4,86
														GCTCACCTCGTGTTC	113	12,61
dnaE	CC1926	GAACAAAACAAGAAA	141	5,44					SMc01375	GAACACGCAGTAAAC	102	7,83	RPA2924	GTTCTTGTTATGTTC	292	1,71
		GAACAAAACCGGAAT	159	5,74						GAACAGTAGCGGAAA	248	9,66		GAACGAAAGAAGAAC	309	2,67
		GAACATCTTGTGAAC	175	6,59						GAACAAAAAGGGAAC	259	0,00				
										GAACGGAAATAGAAC	270	4,15				
dinP	CC2466	-			mlr1877	*GTACATGTTATGTTC*	+9	7,42	SMc01373	*GTTCCGGATATGATC*	+6	6,55	RPA3135	GTTCCGCGGCCGATC	146	12,59
						GTTCTCTATTCGTTC	+20	3,73		GTTCAACATTTGTTC	5	3,67				
					mlr0866	GTTCTCTTTATGTTC	8	2,59								
sulA1	CC3213	GAACAAAAGTGGAAC	66	1,05	mlr4426	GAACATAACAGGAAC	25	1,98	SMc03790	GAACATAACATGAAC	+6	2,65	RPA1801	GAACATATCATGAAC	62	3,87
						GAACGCTGCCCGAGC	66	9,32								
sulA2									SMa0888	GAACAAATACAGAAC	+33	2,08				
sulA3																
parE	CC1974	-			mlr0901	GAACGTAACAGGAAC	73	3,55	SMc01018	ATTCGCCTTTTGTTC	118	5,49	RPA2486	GCACGACATCAGAAC	15	9,39
										GAACGCTTCATGGAC	+6	11,56				
ispE	CC1330	GTTCTTGTTATGTTC	25	1,71	mll7417	GTTCTTGCTTTGTTC	160	2,05	SMc00866	GTTCTTGATTTGTTC	65	0,72	RPA1033	GTTCTTGATTTGTTC	24	0,72
SMc00865									SMc00865	GAACAAATCAAGAAC	101	1,93	RPA1032	GAACAAATCAAGAAC	143	1,93
comM	CC0140	GTTCGTTTTTCGTTC	64	2,84	mll4733	GTTCTCGTAATGTTC	91	4,01	SMc00420	GTTCTATCATTGTTC	139	5,29	RPA0318	GTTCACTTAATGTTC	52	5,82
		TTTCCCGAAACGTTC	75	10,38												
yigN	CC0271	GAACGGAGCATGAAC	+4	6,52	mlr4857	GTTCACGTTTTGATC	2	5,10	SMc01102	GTTCTCGTTTTGATC	35	3,68	RPA0620	GTTCTATCTTCGTTC	*4*	5,61
ppdK	CC1471	-			mlr7532	GTTCCCTTTATGTTC	199	2,91	SMc00025	-			RPA1051	-		
						GGACCAGGGTGGAAC	49	11,14								
dnrV	CC3424	GAACAATTCAGGAAC	20	3,35	mll7529	GAACATAAAGGGAAC	14	1,32	SMc04266	-			RPA1664	GAACACATTGAGCAC	191	7,25
						GTTCCACCCTGGTCC	163	12,47					RPA0902	GTTCAGGGTTCGTGC	81	8,30
recG	CC1437	CAACCCGTGATGAAC	55	12,84	mlr0830	GATCTTTTGCAGAAC	337	14,68	SMc00228	-			RPA2662	GAACTGGCCGAGAAT	249	14,57

Table 3. Distribution of genes with conserved regulatory motifs in the Alpha Proteobacteria. *Name* indicates annotation loci names, whilst *d* denotes distance to the ORF start codon; a + symbol preceding the distance designates intragenic motifs. Shaded rows correspond to regulon core genes. A – symbol indicates that no significant motifs were detected for that gene and species. *Rickettsia prowazekii* and *Rickettsia conorii*, for which no significant results were found, are not included.

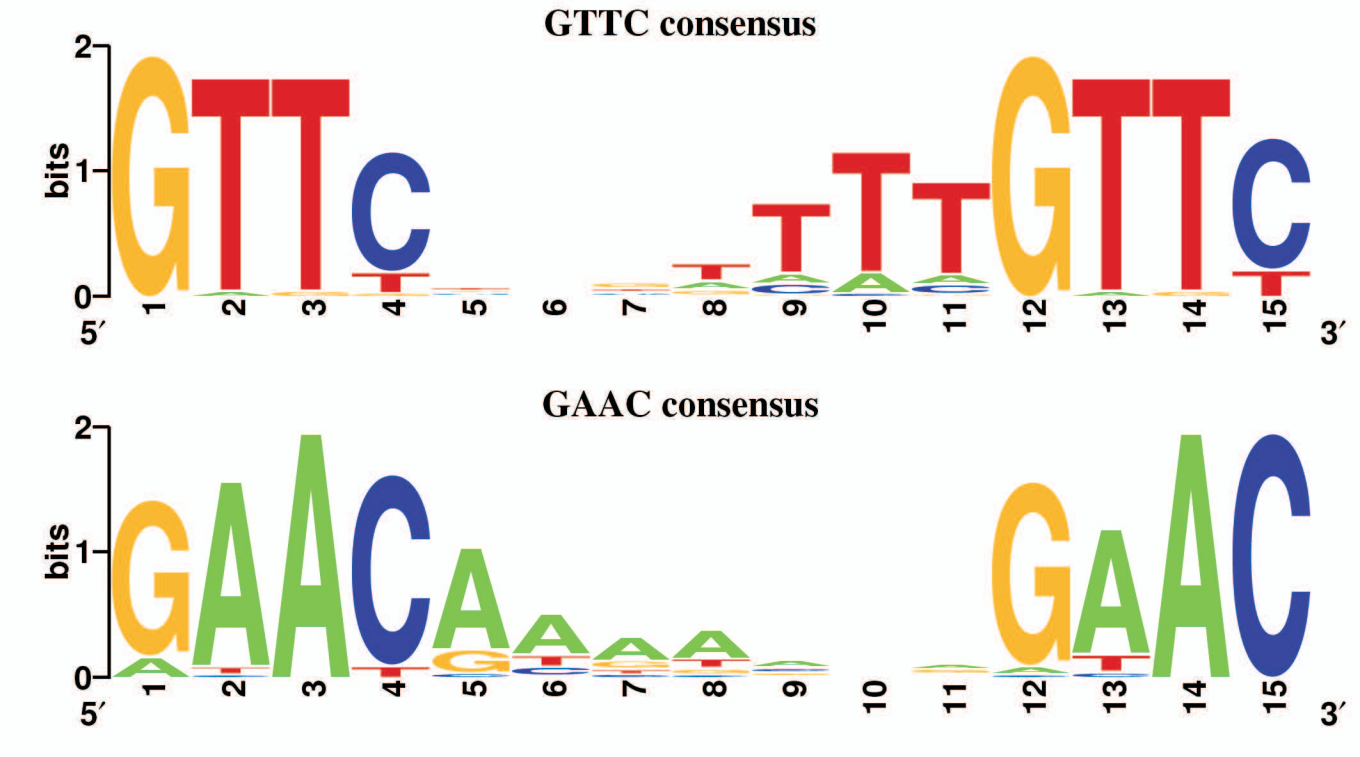
LEGENDS OF FIGURES

Figure 1. Interspecies consensus sequences for Alpha Proteobacteria LexA-binding sites derived from the preliminary analysis and available experimental data (Table 1). Sequence logos were produced using the WebLogo service at http://weblogo.berkeley.edu (44).

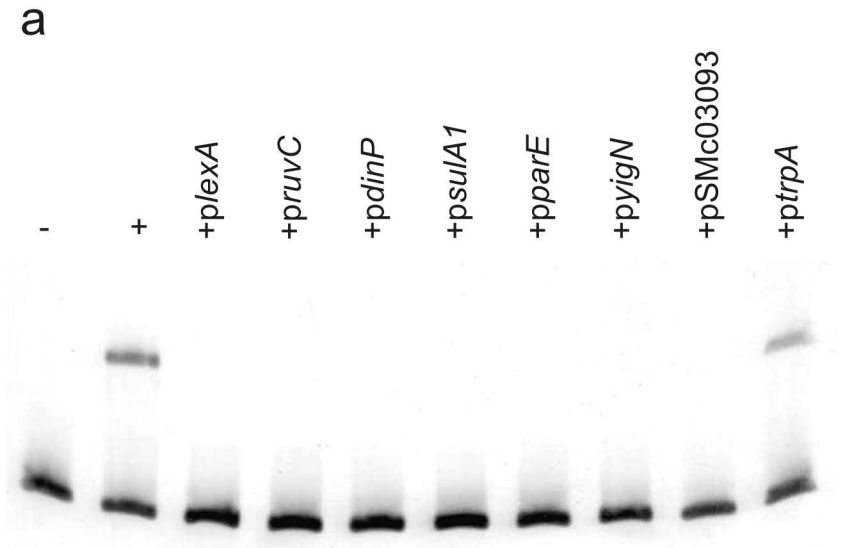
Figure 2. (a) Electrophoretic mobility of the *S. meliloti lexA* promoter in the presence of 80 nM of *S. meliloti* LexA protein and a 300-fold molar excess of unlabeled fragments comprising about 400 bp of the upstream regions of the genes *ruvC*, *dinP*, *sulA1*, *parE*, *yigN* and SMc03093. As a positive control, the effect of unlabelled *lexA* promoter on the mobility of the labeled *lexA* fragment in the presence of the same amount of LexA protein is presented. The mobility of the *lexA* promoter either in the absence of any additional DNA but incubated with LexA protein (+) or in the absence (-) of purified LexA protein is also shown. The *trpA* gene promoter (rightmost lane) was used as a negative control for unspecific binding. (b) Expression of these genes in the presence of mitomycin C at 20 μg / ml. The induction factor (IF) displayed in the rightmost column was computed for each gene as the ratio of relative mRNA concentration in cells treated with mitomycin C to that of untreated ones. The relative mRNA concentration for each gene is normalized to that of the *S. meliloti trpA* gene. Values were calculated 4 h after the addition of mitomycin C. In each case, the mean value from three independent experiments (each in triplicate) is shown, and the standard error of any value in all experiments was always lower than 10%. *d* denotes distance to the ORF start codon; a + symbol preceding the distance designates intragenic motifs

Figure 3. (a) Electrophoretic mobility of the *S. meliloti lexA* promoter in the presence of 80 nM of *S. meliloti* LexA protein and a 300-fold molar excess of unlabelled fragments comprising about 400 bp of the upstream regions of *dnaE*, *ispE*, *sulA2*, *comM*, SMc00865 and SMc03791 genes. As a positive control, the effect of unlabelled *lexA* promoter on the mobility of the labeled *lexA* fragment in the presence of the same amount of LexA protein is presented. The mobility of the *recA* promoter either in the absence of any additional DNA but incubated with LexA protein (+) or in the absence (-) of purified LexA protein is also shown. The *trpA* gene promoter (rightmost lane) was used as a negative control for unspecific binding. (b) Expression of *dnaE*, *ispE*, *sulA2*, *comM* and *SMc00865* genes in the presence of mitomycin C at 20 μg / ml. The induction factor (IF) displayed in the rightmost column is the ratio, for each gene, of relative mRNA concentration in cells treated with mitomycin C to that of untreated ones. The relative mRNA concentration for each gene is normalized to that of the *S. meliloti trpA* gene. Values were calculated 4 h after the addition

of mitomycin C. In each case, the mean value from three independent experiments (each in triplicate) is shown, and the standard error of any value in all experiments was always lower than 10%. *d* denotes distance to the ORF start codon; a + symbol preceding the distance designates intragenic motifs


GTTC consensus
bits
2
1
0
5′
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
3′
GAAC consensus
bits
2
1
0
5′
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
3′

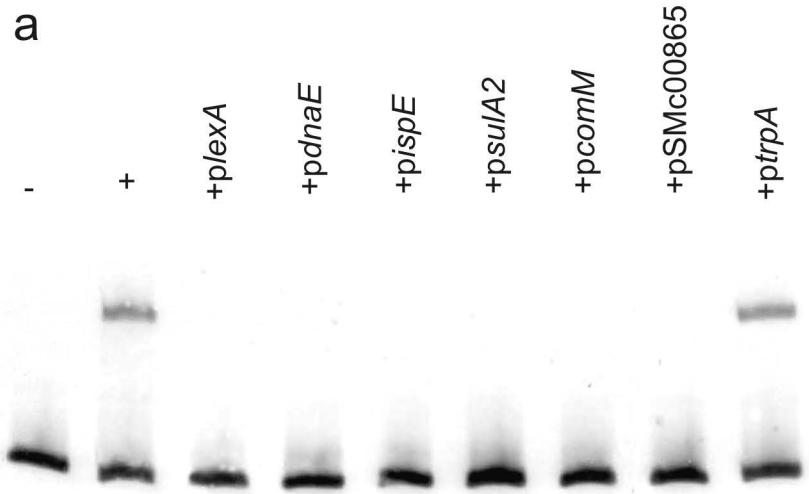
a



b

Gene	Regulatory motif	d	HI	IF
lexA	GTTCTTGATTTGTTT	21	2.297	12
	GAACACATATGGAAC	47	3.113	
ruvC	GTTTTTGTTTTGTTC	44	1.745	5
dinP	*GTTCCGGATATGATC*	+6	6.547	31
	GTTCAACATTTGTTC	5	3.666	
sulA1	GAACATAACATGAAC	+6	2.648	23
parE	ATTCGCCTTTTGTTC	118	5.592	3
yigN	GTTCTCGTTTTGATC	35	3.681	13
SMc03093	GTTCTTGATTTGTTC	148	0.719	39

a



b

Gene	Regulatory motif	d	HI	IF
dnaE	GAACACGCAGTAAAC	102	7.828	2
	GAACAGTAGCGGAAA	248	9.664	
	GAACAAAAAGGGAAC	259	0.001	
	GAACGGAAATAGAAC	270	4.147	
sulA2	GAACAAATACAGAAC	+33	2.085	42
ispE	GTTCTTGATTTGTTC	65	0.719	16
SMc00865	GAACAAATCAAGAAC	101	1.929	4
comM	GTTCTATCATTGTTC	139	5.290	23