Microbiology (2006), **152**, 1109–1118 DOI 10.1099/mic.0.28376-0

Correspondence
Jordi Barbé
jordi.barbe@uab.es

Identification of the *Acidobacterium capsulatum* LexA box reveals a lateral acquisition of the Alphaproteobacteria *lexA* gene

Gerard Mazón,[1] Susana Campoy,[1] Ivan Erill[2] and Jordi Barbé[1,3]

[1]Centre de Recerca en Sanitat Animal (CReSA), Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

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

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

Received 26 July 2005
Revised 16 December 2005
Accepted 19 December 2005

Acidobacterium capsulatum is the most thoroughly studied species of a new bacterial phylogenetic group designated the phylum Acidobacteria. Through a TBLASTN search, the *A. capsulatum* *lexA* gene has been identified, and its product purified. Electrophoretic mobility shift assays have shown that *A. capsulatum* LexA protein binds specifically to the direct repeat GTTCN$_7$GTTC motif. Strikingly, this is also the LexA box of the Alphaproteobacteria, but had not previously been described outside this subclass of the Proteobacteria. In addition, a phylogenetic analysis of the LexA protein clusters together *Acidobacterium* and the Alphaproteobacteria, moving the latter away from their established phylogenetic position as a subclass of the Proteobacteria, and pointing to a lateral gene transfer of the *lexA* gene from the phylum Acidobacteria, or an immediate ancestor, to the Alphaproteobacteria. Lastly, *in vivo* experiments demonstrate that the *A. capsulatum* *recA* gene is DNA-damage inducible, despite the fact that a LexA-binding sequence is not present in its promoter region.

INTRODUCTION

The genetic material of bacteria is often exposed to damaging agents that can compromise its integrity and the viability of the cell. To overcome this problem, many genes are employed in different pathways to detect and repair DNA damage; alternatively, some lesions in the DNA are tolerated, and thus allow a certain degree of mutation. Most of these genes constitute specific networks that act in a coordinated manner, and one such network, present in most Bacteria clades, is the LexA-regulated SOS system.

In *Escherichia coli*, the SOS regulon consists of more than 40 genes, whose functions are involved in DNA replication, DNA repair, mutagenesis and control of the cell cycle (Fernández de Henestrosa *et al.*, 2000; Courcelle *et al.*, 2001). Under normal conditions, the SOS network remains repressed by the product of the *lexA* gene (Walker, 1984). The *E. coli* LexA protein binds specifically to a 16 bp palindrome (CTGTN$_8$ACAG) named the LexA box, located in the promoter region of SOS genes, whose expression is normally down-regulated to a basal level (Little *et al.*, 1981). After DNA-damage-mediated stalling of the replication fork, RecA acquires an active conformation (RecA*) by binding to ssDNA (Sassanfar & Roberts, 1990). RecA* mediates the autohydrolysis of the *E. coli* LexA repressor between residues Ala^{84} and Gly^{85}, resulting in the expression of the SOS genes (Little, 1991; Little *et al.*, 1980). Once DNA damage is effectively repaired, RecA* concentration decreases, LexA ceases to be hydrolysed, and the non-cleaved LexA synthesized *de novo* once again inhibits the transcriptional expression of the SOS network.

The evolutionary history of the SOS system across the domain Bacteria is complex. First of all, and with very few exceptions (such as *Cytophaga*, *Flavobacterium*, *Bacteroides* and Epsilonproteobacteria), the *lexA* gene seems to be present in all bacterial families in which drastic genetic reduction (such as that observed in *Buchnera*, *Rickettsia*, *Aquifex* and *Mycoplasma*, among others) has not taken place. Furthermore, a clear relationship seems to exist between the LexA-binding sequence and the branching order of several bacterial phylogenetic groups from their common ancestor (Mazón *et al.*, 2004b). Therefore, and in agreement with the bacterial genome sequences known so far, several groups can be established. The first comprises all the Gram-positive bacteria and closely related phyla, such as the Cyanobacteria and the green non-sulfur bacteria. All these phyla possess the GAACN$_4$GTTC sequence or a related motif as their specific LexA-binding sequence (Winterling *et al.*, 1998; Fernández

Abbreviations: EMSA, electrophoretic mobility shift assay; LGT, lateral gene transfer.

Supplementary data is available with the online version of this paper.

de Henestrosa *et al.*, 2002; Mazón *et al.*, 2004a). On the other hand, Alphaproteobacteria have a $GTTCN_7GTTC$ direct repeat as their LexA box (Fernández de Henestrosa *et al.*, 1998; Tapias & Barbé, 1999), whereas Beta- and Gammaproteobacteria (Erill *et al.*, 2003) share the same LexA-binding sequence as that found in *E. coli* ($CTGTN_8ACAG$). However, an intermediate phylum (Fibrobacter) between Proteobacteria and Cyanobacteria possesses a LexA box that seems to represent a transition from the Gram-positive one to that seen in *E. coli* (Mazón *et al.*, 2004b).

Several DNA repair genes, including some belonging to the canonical LexA regulon (*recA* and *uvrA*), are usually involved in the resistance of bacteria to acidic pH (Raja *et al.*, 1991; Thompson & Blaser, 1995; Hanna *et al.*, 2001). However, some bacterial species whose environmental growth conditions are very acidic show a constitutive expression of the *recA* gene that is mediated either by the absence of a LexA box on the *recA* promoter (e.g. *Acidobacillus thioferrooxidans*; Ramesar *et al.*, 1989) or by the lack of a *lexA* regulatory gene (e.g. *Helicobacter pylori*; Tomb *et al.*, 1997).

Acidobacterium capsulatum is a Gram-negative, acidophilic, chemo-organotrophic bacterium, containing menaquinone, of which very few isolates have been cultured (Kishimoto *et al.*, 1991; Hiraishi *et al.*, 1995). This micro-organism has been proposed to belong to a new bacterial phylum that branched approximately at the same time as the phylum Fibrobacter, and significantly earlier than the Proteobacteria (Ludwig *et al.*, 1997; Quaiser *et al.*, 2003). Despite its extreme growth conditions, no data exist on the DNA repair system of *A. capsulatum*. For this reason, and taking advantage of the fact that the *A. capsulatum* genome is being sequenced (http://www.tigr.org), here we report the cloning of the *A. capsulatum lexA* gene, the purification of its product, and the characterization of its binding site, in an effort to unravel both the composition of its LexA regulon and the evolution of this gene network across the domain Bacteria.

METHODS

Bacterial strains, plasmids, oligonucleotides and DNA techniques. Bacterial strains and plasmids used in this work are listed in Table 1. *E. coli* was grown in LB medium at 37 °C and antibiotics were added to the cultures at the concentrations reported in Sambrook *et al.* (1992). *A. capsulatum* ATCC 51196 was grown in M-269 medium with $0{\cdot}1\ g\ l^{-1}$ yeast extract (pH 3·5) at 30 °C (Hiraishi *et al.*, 1995). *E. coli* cells were transformed with plasmid DNA as described by Sambrook *et al.* (1992). Oligonucleotide primers for PCR, restriction enzymes, T4 DNA ligase and polymerase, and the 'DIG DNA labelling and detection kit', were from Roche. Genomic DNA of *A. capsulatum* was obtained by standard procedures (Sambrook *et al.*, 1992) from a 10 ml culture grown at 30 °C in M-269. The DNA sequence of all PCR-mutagenized fragments was determined by the dideoxy method (Sanger *et al.*, 1977) on an ALF sequencer (Amersham-Pharmacia).

Identification, cloning and purification of *A. capsulatum* LexA. The *A. capsulatum lexA* gene sequence was identified by performing a TBLASTN search of its unfinished genome at The Institute for Genomic Research (TIGR) (http://tigrblast.tigr.org/ufmg/) using *Fibrobacter succinogenes* LexA protein as a query. The comparison yielded a region with significant homology in part of contig number 83 (Supplementary Table S1), and we obtained this region with 1 kb flanking each side of the hypothetical *lexA* coding region. Primers Up-Acido-LexA (5′-CTTCTCCTGCACGCAAGC-3′) and Dw-Acido-LexA (5′-*GGATCC*TACTTGCGTCTGAAATCG-3′)

Table 1. Bacterial strains and plasmids used in this work

Strain or plasmid	Relevant features	Source or reference
Acidobacterium capsulatum		
ATCC 51196	Wild-type	ATCC*
Escherichia coli		
DH5α	*supE4* Δ*lacU169* (Ø80 *lacZ*ΔM15) *hsdR17 recA1 endA1 gyrA96 thi*-1 *relA1*	Clontech
BL21(DE3) Codon Plus	F^- *ompT* $hsdS_B(r_B\text{-}r_B)$ *dcm gal* λ(DE3) *endA* Hte [*argU ileY leuW* Cam^R]	Stratagene
Plasmids		
pGemT	PCR cloning vector; Ap^R	Promega
pET15b	Overexpression vector carrying an N-terminal tag containing six histidine residues; Ap^R	Novagen
pUA1066	pGemT derivative carrying an 813 bp PCR fragment amplified with Up LexA Ac and Dw LexA Ac primers containing the promoter region and the coding region of *A. capsulatum lexA*	This work
pUA1067	pGemT derivative carrying a 625 bp PCR fragment amplified with Nde LexA Ac and Dw LexA Ac primers containing the *A. capsulatum lexA* coding region	This work
pUA1068	A pET15b derivative carrying a *Nde*I–*Bam*HI 625 bp fragment containing the *A. capsulatum lexA* coding region	This work

*American Type Culture Collection.

(italic type shows restriction sites) were designed using this sequence, and were employed to amplify the entire *A. capsulatum lexA* gene and its promoter region by PCR. The resulting 813 bp fragment was cloned into a pGEM-T vector, resulting in pUA1066. This plasmid was then sequenced to check for the introduction of any changes in the DNA sequence during PCR and to confirm the database DNA sequence. *A. capsulatum* LexA contains all the conserved residues involved in the repressor autocleavage (Ala, Gly, Ser and Lys), as deduced from a CLUSTALW alignment of different LexA proteins (Fig. 1).

A. capsulatum LexA protein was purified using the TALON Purification kit (Stratagene). Briefly, the coding sequence of the *A. capsulatum lexA* gene was amplified and cloned using primers *Nde*I-Acido-LexA (5′-*CATATG*GCTGTGACCAAGCGTC-3′) and Dw-Acido-LexA, which contained *Nde*I and *Bam*HI restriction sites, respectively. The resulting pUA1067 plasmid was then digested with *Nde*I/*Bam*HI, and the 625 bp DNA fragment was cloned into pET15b (Novagene). The latter plasmid was transformed into *E. coli* BL21 (DE) Codon plus cells (Stratagene) to overproduce the LexA protein, which was purified by Co^{2+} affinity chromatography, as described previously (Campoy *et al.*, 2002). *Rhodobacter sphaeroides* LexA protein, also used in this work, had been previously purified (Tapias *et al.*, 2002).

Mobility shift assays. LexA–DNA complexes were detected by electrophoretic mobility shift assays (EMSAs), using purified LexA proteins. Typical 20 μl reactions containing 10 ng DIG-DNA labelled probe and 20 nM final concentration of either pure *A. capsulatum* or pure *R. sphaeroides* LexA were incubated in binding buffer (5 mM Tris/HCl, 5 mM $MgCl_2$, 25 mM NaCl, 5 %, v/v, glycerol, 1 μg bulk carrier DNA and 50 μg BSA ml^{-1}). After 30 min incubation, the reaction mixture was loaded onto a 6 % non-denaturing Tris/glycine/polyacrylamide gel. DNA–protein complexes were separated at 100 V for 1 h, and then transferred to a Biodine B membrane (Pall Gelman Laboratory), followed by the detection of DIG-labelled DNA–protein complexes, following the manufacturer's protocol (Roche).

RT-PCR analysis. Mitomycin C-mediated induction of several genes was analysed by real-time quantitative RT-PCR of total *A. capsulatum* RNA with LightCycler (Roche), as described previously (Campoy *et al.*, 2002; Bustin, 2002), and using the appropriate primers. The Titan One Tube RT-PCR system (Roche) was used, as described by Campoy *et al.* (2002), to perform RT-PCR experiments and to determine the transcriptional organization of the *imuA-imuB-dnaE* gene cassette.

Phylogenetic analysis. Sequence data of *F. succinogenes*, *Myxococcus xanthus* and *A. capsulatum* were obtained from TIGR through their website at http://www.tigr.org. Protein sequences for all other organisms were acquired from the Microbial Genome Database for Comparative Analysis website (http://mbgd.genome.ad.jp/) and the US Department of Energy Joint Genome Institute website (http://genome.jgi-psf.org/mic_home.html). All proteins used for phylogenetic analyses are listed in Supplementary Table S1, where their locus, accession numbers and micro-organism abbreviations are also given. Alignments of protein sequences were carried out using a combined procedure to improve alignment quality. Protein sequences were first aligned through CLUSTALW version 1.83 (Thompson *et al.*, 1994), using default (10), 25 and 5 gap-opening penalties (GOPs) for the multiple alignment stage, and generating three different

```
                                                    α3
Aca   -MAVTKRQKEVLDFITGFVQRNGYSPSFEEIARGLNLKSLATVHKHITNLQNKGLLARAH   59
Sme   --MLTRKQQELLLFIHERMKESGVPPSFDEMKDALDLASKSGIHRLITALEERGFIRRLP   58
Rca   --MLTHKQLELLLDFIQKRMARDGVPPSFDEMKDALDLRSKSGIHRLITALEARGFIRRMA   58
Bsu   MTKLSKRQLDILRFIKAEVKSKGYPPSVREIGEAVGLASSSTVHGHLARLETKGLIRRDP   60
Fsu   ---MTARQEEIYEYIKKYSKENHMPPTVREIGNHFDISSTNGVRSILAALIKKGYINRSP   57
Eco   MKALTARQQEVFDLIRDHISQTGMPPTRAEIAQRLGFRSPNAAEEHLKALARKGVIEIVS   60

Aca   NRSRSLDVLPP---RSRGRSVDR---------------------------------LPLM   83
Sme   NRARALEVIKLPEAYAGASQVRRGFSPSVIEGSLGKPAAPPPAPKPAPPAEAASVAVPVM  118
Rca   HRARALEIVKLPESMERAAEAERGLQTQPAFTPTLIAGSRTEPPRGALPVAAAALDIPMM  118
Bsu   TKPRAIEILDEEVDIPQSQVVN----------------------------------VPVI   86
Fsu   RLSRGIEILSDDKESSKEVASNT-------------------------------IEIPIV   86
Eco   GASRGIRLLQEEEEG-----------------------------------------LPLV   79

Aca   GRIAAGRPVEAAEN---AESISLGDIIGNREVFALQVHGESMRDEHIVDGDYVLVERTNT  140
Sme   GRIAAGVPISAIQNNMHDISVPVEMI-GSGEHYALEIKGDSMIEAGILDGDTVIIRNGST  177
Rca   GRIAAGVPIEAIAEVAHHVTVPGAMLSGRGEHYALEVKGDSMIQAGINDGDIVVIRSQPT  178
Bsu   GKVTAGSPITAVENIEEYFPLPDRMVPPDEHVFMLEIMGDSMIDAGILDKDYVIVKQQNT  146
Fsu   GRVAAGTPILAVQNLEGTVTIDRDFLACRSDVFALRVKGDSMINAGIFDGDLIFARQQKT  146
Eco   GRVAAGEPLLAQQHIEGHYQVDPSLFKPNAD-FLLRVSGMSMKDIGIMDGDLLAVHKTQD  138

Aca   ARQGEIIVALVRGAETTLKRFYLEGSMVRLQPSNAEMNPIIVP--AAQVAIQGRVLGMLR  198
Sme   ASPGDIVVALIDDEEATLKRFRRRGASIALEAANPAYETRIFG--PDRVKIKGRLVGLIR  235
Rca   AENGDIVVALVEDLEATLKRYYRRGGMIALEAANPAYETRLLR--EDQVKVQGRLVGLIR  236
Bsu   ANNGEIVVAMTEDDEATVKRFYKEDTHIRLQPENPTMEPIILQ--N--VSILGKVIGVFR  202
Fsu   ADLGEIVVAQIDN-EATVKYYHPSADHVELRPANPKYKPIIVNNRKD-FSIAGRVIGVMR  204
Eco   VRNGQVVVARIDD-EVTVKRLKKQGNKVELLPENSEFKPIVVDLRQQSFTIEGLAVGVIR  197

Aca   KY---  200
Sme   RYH--  238
Rca   SY---  238
Bsu   TVH--  205
Fsu   KVN--  207
Eco   NGDWL  202
```

Fig. 1. CLUSTALW alignment comparison of the LexA proteins from *A. capsulatum* (*Aca*), *Sinorhizobium meliloti* (*Sme*), *Rhodobacter capsulatus* (*Rca*), *Bacillus subtilis* (*Bsu*), *Fibrobacter succinogenes* (*Fsu*) and *Escherichia coli* (*Eco*). Dark shading shows identical conserved amino acids; light shading indicates similar conserved residues. Arrows indicate the Ala, Gly, Ser and Lys residues involved in the LexA autocatalytic cleavage. The residues of the α3 helix, the DNA binding domain, are overlined.

alignments. These three alignments, together with a local alignment generated by the T-COFFEE lalign method, were integrated as libraries into T-COFFEE version 1.37 (Notredame *et al.*, 2000) for optimization. The optimized alignment was then visually inspected with BioEdit version 5.0.9 (Hall, 1999) and submitted to Gblocks version 0.91b (Castresana, 2000) with the half-gaps setting and, otherwise, default parameters, in order to select conserved positions and discard poorly aligned and phylogenetically unreliable information.

Phylogenetic analyses of the refined alignments were carried out using MrBayes version 3.1.1 (Ronquist & Huelsenbeck, 2003), applying a mixed four-category Gamma distributed rate model plus proportion of invariable sites model (invgamma). MrBayes Metropolis-Coupled Markov Chain Monte Carlo runs were carried out with four independent chains for 10^6 generations. The resulting phylogenetic trees, which are the product of four independent MrBayes runs, were plotted with TREEVIEW version 1.6.6 (Page, 1996). In all cases, only branch support values over 0·85 are shown.

RESULTS

Identification of the *A. capsulatum* LexA-binding site

EMSAs with the purified *A. capsulatum* LexA were used to determine the ability of this protein to bind to its own promoter. The addition of increasing concentrations of LexA to a fragment extending from −149 to +112 of the *A. capsulatum lexA* gene promoter (with respect to its predicted translational starting point) produced one retardation band whose intensity was directly related to the amount of protein used (Fig. 2b). To locate more precisely the LexA-binding region, the *A. capsulatum lexA* promoter was divided into two fragments (designated LexA1 and LexA2), which were obtained by PCR amplification with suitable oligonucleotides. The two fragments were DIG end-labelled and then used as probes in an EMSA (Fig. 2a). A stable DNA–protein complex was observed when both fragments were incubated in the presence of purified *A. capsulatum* LexA (Fig. 2c). These data indicate that the region to which the LexA protein binds was located downstream of position −35 of the *lexA* promoter.

A close inspection of this region revealed the presence of an Alphaproteobacteria-like LexA-binding site (GTTCN$_7$GTTC), extending from positions −29 bp to −15 bp with respect to the *lexA* translational start codon (Fig. 2a). To test whether this sequence was actually the *A. capsulatum* LexA box, an EMSA experiment with a LexA2-derivative fragment, in

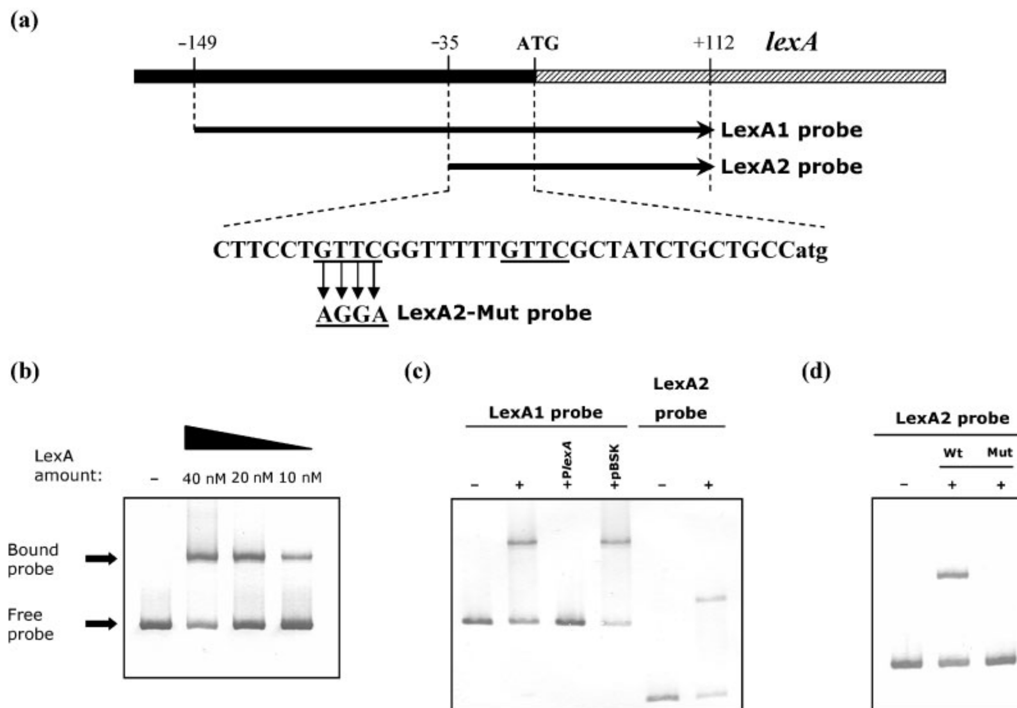


Fig. 2. (a) Fragments amplified from the *A. capsulatum lexA* promoter region and used in electrophoretic mobility shift assays (EMSAs). In all cases, positions are indicated with reference to the translational start point of the *lexA* gene. (b) Electrophoretic mobility of the LexA1 probe in the presence of decreasing concentrations of LexA protein. (c) Effect of a 500-fold molar excess of unlabelled *A. capsulatum lexA* promoter (P*lexA*) and pBSK(+) plasmid DNA (pBSK) in the migration of the LexA1 probe in the presence of 20 nM *A. capsulatum* LexA protein. Migration of LexA2 probe in the presence (+) or absence (−) of *A. capsulatum* LexA is also shown. (d) Effect of a tetranucleotide substitution (LexA2-Mut) in the migration of the LexA2 probe (Mut) in the presence of 20 nM pure *A. capsulatum* LexA protein. The mobility of the wild-type LexA2 probe (Wt) in the presence (+) and absence (−) of LexA protein is also shown.

which the first GTTC motif was substituted by an AGGT tetranucleotide, was performed. Fig. 2(d) shows that the GTTC motif was needed by *A. capsulatum* LexA to successfully bind its promoter. Furthermore, *R. sphaeroides* LexA protein was also able to bind the wild-type LexA2 fragment, but not its mutant derivative (Fig. 3a). Likewise, the *A. capsulatum* LexA protein was able to bind the wild-type *R. sphaeroides recA* promoter but not a mutant derivative with a modified LexA Box (Fig. 3b). These experiments clearly demonstrate that *R. sphaeroides* and *A. capsulatum* recognize the same LexA-binding sequence. In this respect, it is worth noting that the same LexA box ($GTTCN_7GTTC$) is present in the *lexA* gene promoter region of *Solibacter usitatus* sp. Ellin6076, another member of the phylum Acidobacteria whose sequence is being completed by the Joint Genome Institute (http://www.jgi.doe.gov).

Analyses of the LexA-regulon gene core

Comparative analyses of the SOS system in different subclasses of the Proteobacteria (Alpha, Beta and Gamma) indicate that a common set of genes (*lexA*, *recA*, *ssb*, *uvrA* and *ruvA*) is directly repressed by LexA in all these subclasses, and therefore constitutes the canonical gene composition of the SOS regulon in this phylum (Erill *et al.*, 2003, 2004). Some of these genes (*lexA*, *recA* and *ruvA*) are also regulated by LexA in *Bacillus subtilis* (Dubnau & Lovett, 2002). Furthermore, it has recently been reported that a gene (*dnaE2*) that encodes a second alpha subunit of DNA polymerase III, and which is found either isolated in the chromosome (*Mycobacterium tuberculosis*) or belonging to a multiple gene cassette (many Beta- and Gammaproteobacteria species), is also under LexA control (Boshoff *et al.*, 2003; Abella *et al.*, 2004; Galhardo *et al.*, 2005). To analyse the constitution of the *A. capsulatum* LexA regulon, TBLASTN was employed to search the *A. capsulatum* genomic database for *recA*, *dinP*, *uvrA*, *ruvAB*, *ssb* and *dnaE2*, using their respective homologues in the Alphaproteobacteria. It is worth noting that, besides all the searched proteobacterial or *B. subtilis lexA*-regulated genes (*recA*, *dinP*, *uvrA*, *ruvAB* and *ssb*), two independent copies (*iid1* and *iid2*) of the multiple gene cassette constituted by the *imuA-imuB-dnaE2* genes were also detected. Furthermore, the *iid1* and *iid2* cassettes both constitute single transcriptional units when analysed by RT-PCR (Fig. 4).

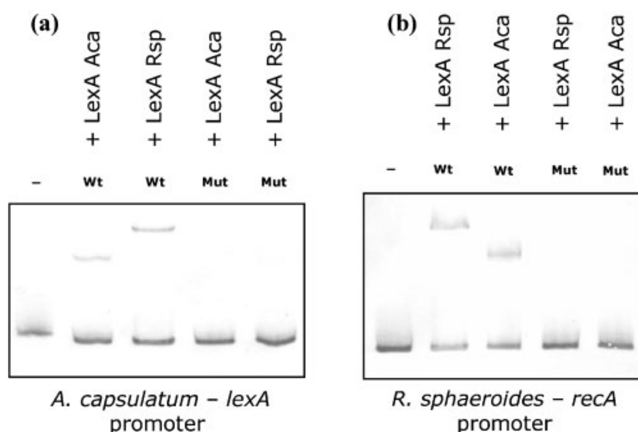


Fig. 3. (a) Migration of both *A. capsulatum* LexA2 (Wt) and LexA2-Mut (Mut) probes in the presence of 20 nM LexA protein of either *A. capsulatum* (LexA Aca) or *R. sphaeroides* (LexA Rsp). (b) Migration of both *R. sphaeroides recA* wild-type (Wt) promoter and a mutant derivative (Fernández de Henestrosa *et al.*, 1998) lacking its LexA box (Mut) in the presence of 20 nM LexA protein of either *A. capsulatum* (LexA Aca) or *R. sphaeroides* (LexA Rsp).

After the identification of these genes and operons (*iid1*, *iid2*, *ruvA*, *dinP*, *recA* and *uvrA*) in the *A. capsulatum* database, DNA fragments containing their upstream regions were isolated by PCR using suitable oligonucleotides, and the ability of these fragments to bind the *A. capsulatum* LexA protein was tested. Accordingly, competitive EMSAs were carried out with the LexA2 DIG-labelled probe and a 500-fold excess of non-labelled promoter fragments of all these genes. Fig. 5 shows that only *iid1*, *iid2* and *dinP* upstream regions were able to bind the *A. capsulatum* LexA protein. It must be noted that a sequence search revealed that there are no ORFs immediately upstream of these genes, indicating that each gene is the first one of a transcriptional unit (data not shown). In accordance with these results, *A. capsulatum* LexA-binding sites were only found in *iid1*, *iid2* and *dinP* promoter regions (data not shown).

To further characterize the behaviour of the DNA-repair-related *A. capsulatum* genes identified, the effect of mitomycin C on their expression was analysed. Table 2 shows that, as expected, *lexA*, *iid1* and *iid2* transcriptional units are induced by DNA damage. It is worth noting also that, among the canonical proteobacterial LexA-regulated genes present in *A. capsulatum*, only *recA* is DNA-damage inducible in this organism, despite the fact LexA does not bind to its promoter region (Fig. 5).

Phylogenetic comparison of the *A. capsulatum lexA* gene

It has been shown that the phylum Acidobacteria branched long before the Proteobacteria on the evolutionary tree (Ludwig *et al.*, 1997), and the same holds true for the phylum Fibrobacter (Griffiths & Gupta, 2001). Nevertheless, *F. succinogenes* possesses a LexA-binding sequence that seems to represent a transition between that of Gram-positive bacteria and that of Beta- and Gammaproteobacteria (Mazón *et al.*, 2004b). In contrast, the *A. capsulatum* LexA box seems clearly related to that of the Alphaproteobacteria, which branched later than Deltaproteobacteria, a subclass in which *F. succinogenes*-related LexA recognition motifs can be observed (Campoy *et al.*, 2003; Mazón *et al.*, 2004b). This fact strongly suggested the possibility of a lateral gene transfer (LGT) event concerning the *lexA* gene between Acidobacteria and Alphaproteobacteria.

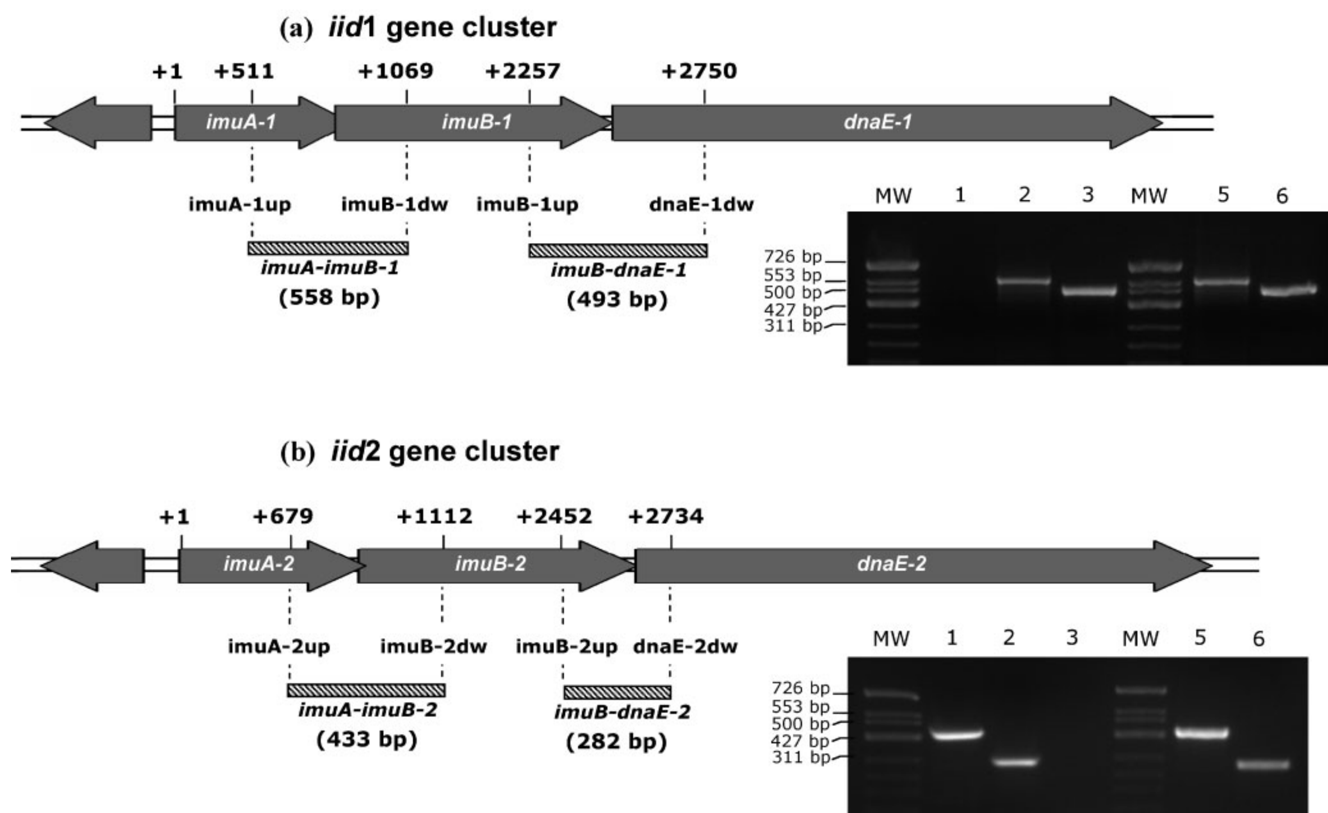


Fig. 4. Structural arrangement of both copies (*iid1* and *iid2*) of the *A. capsulatum imuA-imuB-dnaE* transcriptional unit. Positions are indicated with reference to the translational start point of the *imuA* gene of each operon, and mark the 5′ end of the primers designed to amplify the *imuA-imuB* and *imuB-dnaE* fragments by RT-PCR. (a) RT-PCR analysis of *iid1* with imuA-1up and imuB-1dw primers (lane 2) and imuB-1up and dnaE-1dw primers (lane 3); as a control, the same reactions were done using DNA (lanes 5 and 6) or RNA in a PCR reaction with imuA-1up and imuB-1dw primers (lane 1). (b) RT-PCR analysis of *iid2* with imuA-2up and imuB-2dw primers (lane 1) and imuB-2up and dnaE-2dw primers (lane 2); as a control the same reactions were done using DNA (lanes 5 and 6) or RNA in a PCR reaction with imuA-2up and imuB-2dw primers (lane 3). The *Hinf*I-digested DNA of ϕX174 used as a molecular weight (MW) marker is also shown.

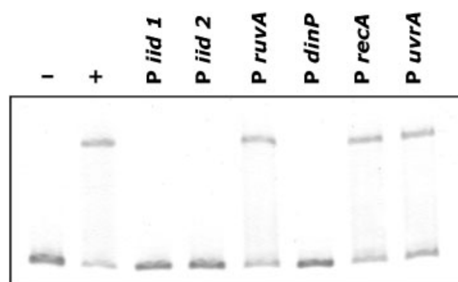


Fig. 5. EMSA experiments using the LexA2 probe incubated with purified LexA protein of *A. capsulatum* in the presence of different DNA competitors. The LexA2 probe was incubated in the presence (+) or absence (–) of 20 nM purified *A. capsulatum* LexA protein and, at the same time, in the presence of 20 nM LexA and a 500-fold molar excess of unlabelled fragments containing promoter regions from *recA*, *uvrA*, *ruvA* and *dinP* genes, and the *iid*1 and *iid*2 transcriptional units.

Accordingly, phylogenetic analyses using the LexA protein reveal a strong relationship between Acidobacteria and Alphaproteobacteria. As can be seen in Fig. 6, both types of LexA protein cluster together with a high support value. Conversely, when other proteins belonging to the Alphaproteobacteria SOS regulon (such as RecA, UvrA, Ssb and

Table 2. Behaviour of several DNA-repair-related *A. capsulatum* genes towards mitomycin exposure

Gene	LexA binding	Induction factor
lexA	+	10·84
iid1	+	48·92
iid2	+	24·29
recA	–	4·43
uvrA	–	1·03
ruvA	–	1·05
ssb	–	1·08

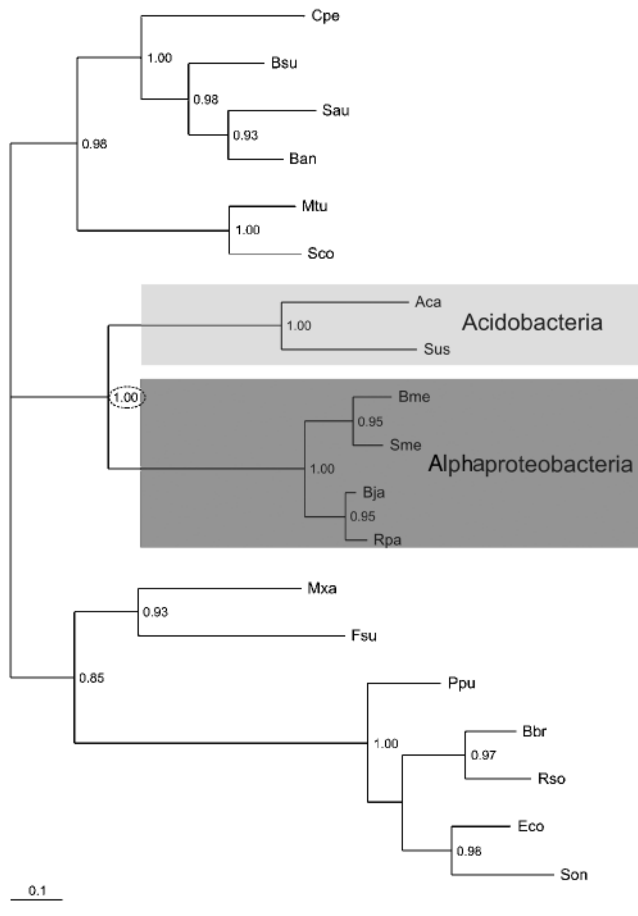


Fig. 6. LexA protein sequence tree. The light-shaded area corresponds to the phylum Acidobacteria, while dark shading designates the Alphaproteobacteria subclass. Only branch support values over 0·85 are shown. The relevant branch support value is encircled. The scale bar indicates 0·1 expected changes per site. Name abbreviations are as follows: Aca, *Acidobacterium capsulatum*; Ban, *Bacillus anthracis*; Bbr, *Bordetella bronchiseptica*; Bja, *Bradyrhizobium japonicum*; Bme, *Brucella melitensis*; Bsu, *Bacillus subtilis*; Cpe, *Clostridium perfringens*; Eco, *Escherichia coli*; Fsu, *Fibrobacter succinogenes*; Mtu, *Mycobacterium tuberculosis*; Mxa, *Myxococcus xanthus*; Ppu, *Pseudomonas putida*; Rpa, *Rhodopseudomonas palustris*; Rso, *Ralstonia solanacearum*; Sau, *Staphylococcus aureus*; Sco, *Streptomyces coelicolor*; Sme, *Sinorhizobium meliloti*; Son, *Shewanella oneidensis*; Sus, *Solibacter usitatus* sp. Ellin6076. Additional information for each LexA is provided in Supplementary Table S1.

UvrD) are used for phylogenetic analysis, the placement of the Alphaproteobacteria is in concordance with the established phylogenetic location of that phylum (Ludwig *et al.*, 1997), and is also supported by high support values (Fig. 7).

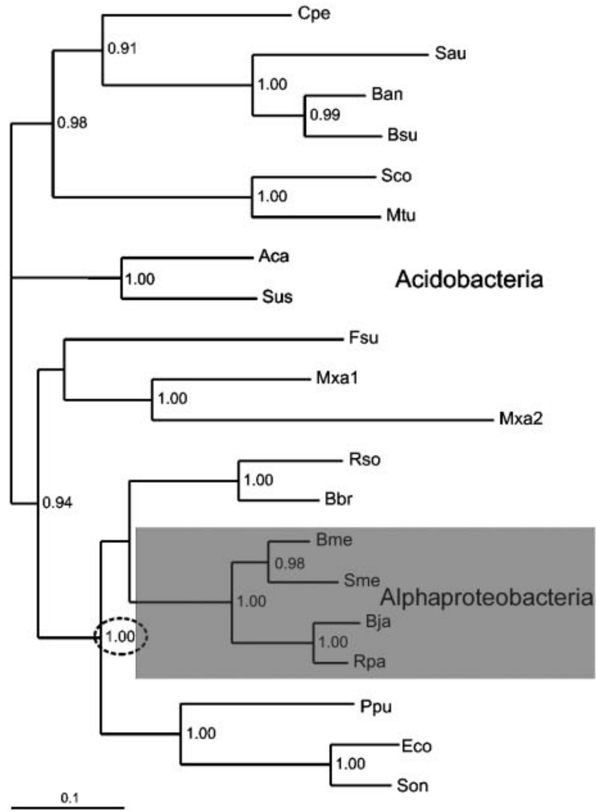
DISCUSSION

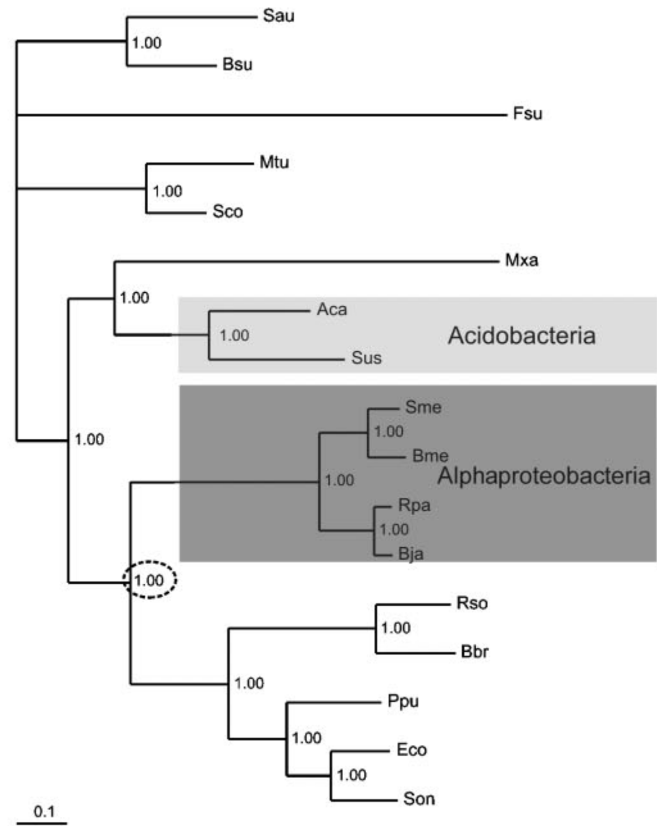
In the last few years, different LexA-binding motifs have been described for a number of bacterial clades and, as is to be expected, most of these motifs share a substantial number of features. In particular, almost all the LexA boxes described so far present a palindromic dyad–spacer–dyad structure, and are monophyletic to the clade in which they have been described. Thus, a GAACN$_4$GTTC palindrome is the consensus sequence for Gram-positive bacteria (Winterling *et al.*, 1998), and this motif, with slight variations, extends also to green non-sulfur bacteria (Fernández de Henestrosa *et al.*, 2002) and Cyanobacteria (Mazón *et al.*, 2004a). The LexA-binding motif is overspecified in terms of information content (LexA boxes are apparently too long for the number of LexA sites in the genome), and this may be the reason behind the vast diversification observed in LexA-binding motifs, which has led to the proliferation of seemingly dissimilar LexA boxes, such as the aforementioned Gram-positive one and that observed in *E. coli* (CTGTN$_8$ACAG). In spite of this, several major features can be identified among these divergent motifs, and we have previously provided experimental support for a basic evolutionary history of LexA-binding motifs based on a vertical inheritance model (Mazón *et al.*, 2004b).

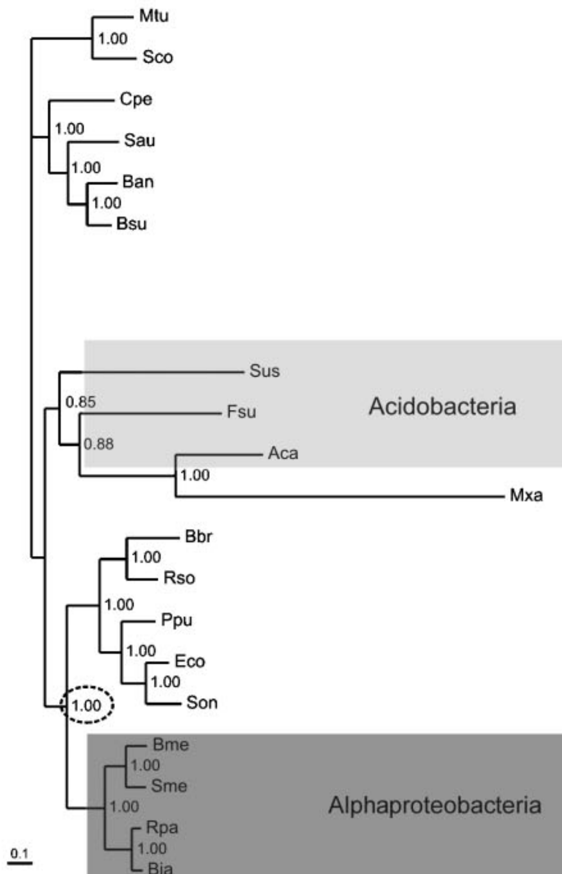
Among the different LexA motifs described so far, the Alphaproteobacteria LexA box is perhaps the most divergent and the most intriguing. First described in *R. sphaeroides* (Fernández de Henestrosa *et al.*, 1998), the Alphaproteobacteria LexA box is monophyletic for this proteobacterial subclass (Erill *et al.*, 2004), and presents a direct-repeat structure that, up to now, had not been described for any other bacterial clade. This fundamental departure from the standard palindromic structure of all other known LexA-binding motifs, together with a highly divergent protein sequence, have made it difficult to reconstruct the precise evolutionary history of the Alphaproteobacteria LexA, but we have previously shown that its most probable origin lies in an evolutionary pathway different from the one that gave rise to the Gamma- and Betaproteobacteria LexA proteins (Mazón *et al.*, 2004b).

The identification here of an Alphaproteobacteria-like LexA box in the Acidobacteria, a lineage that is supposed to branch deeply from either the Chlamydia/Planctomyces or the Gram-positive lines (Hiraishi *et al.*, 1995), thus represents a key piece of evidence in this respect and gives rise to several questions. On the one hand, the identification of the *A. capsulatum* LexA box demonstrates that the GTTCN$_7$GTTC LexA-binding motif existed well before the Alphaproteobacteria emerged. Based on the tendency towards diversification observed in the LexA-binding sequence, it seems very unlikely that the same motif, and precisely such a divergent motif, should have arisen independently in two different phyla. Moreover, the ability reported here of Alphaproteobacteria and *Acidobacterium* LexA proteins to effectively bind one another's promoters is further strong evidence of a common origin for both proteins.

Taking into account that no motif similar to that of the Alphaproteobacteria has been identified so far in any of the studied phylogenetic groups that are intermediate between


(a)RecA
Cpe
0.91
Sau
1.00
Ban
0.99
Bsu
0.98
Sco
1.00
Mtu
Aca
1.00
Sus
Acidobacteria
Fsu
Mxa1
1.00
Mxa2
Rso
1.00
Bbr
0.94
Bme
0.98
Sme
1.00
Alphaproteobacteria
Bja
1.00
Rpa
1.00
Ppu
1.00
Eco
1.00
Son
0.1


(b)UvrD
Sau
1.00
Bsu
Fsu
Mtu
1.00
Sco
Mxa
1.00
Aca
1.00
Sus
Acidobacteria
Sme
1.00
Bme
1.00
Alphaproteobacteria
Rpa
1.00
Bja
1.00
1.00
Rso
1.00
Bbr
1.00
Ppu
1.00
Eco
1.00
Son
0.1


(c)UvrA
Mtu
1.00
Sco
Cpe
1.00
Sau
1.00
Ban
1.00
Bsu
Sus
0.85
Fsu
Acidobacteria
0.88
Aca
1.00
Mxa
Bbr
1.00
Rso
1.00
Ppu
1.00
Eco
1.00
Son
1.00
Bme
1.00
Sme
1.00
Alphaproteobacteria
Rpa
1.00
Bja
0.1

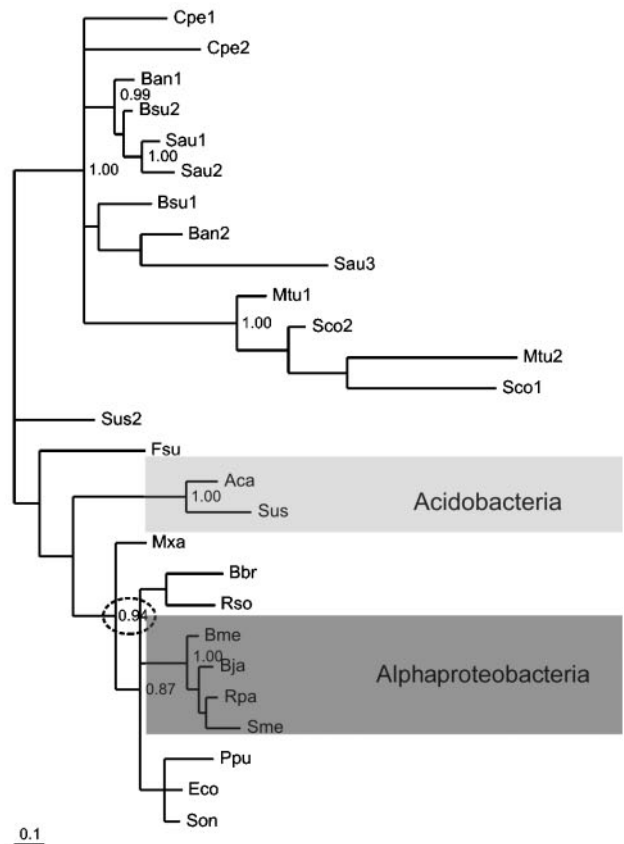

(d)Ssb
Cpe1
Cpe2
Ban1
0.99
Bsu2
Sau1
1.00
Sau2
1.00
Bsu1
Ban2
Sau3
Mtu1
1.00
Sco2
Mtu2
Sco1
Sus2
Fsu
Aca
1.00
Sus
Acidobacteria
Mxa
Bbr
Rso
0.94
Bme
1.00
Bja
0.87
Alphaproteobacteria
Rpa
Sme
Ppu
Eco
Son
0.1

Fig. 7. RecA (a), UvrD (b), UvrA (c) and Ssb (d) protein sequence trees. The light-shaded area corresponds to the phylum Acidobacteria, while dark shading designates the Alphaproteobacteria subclass. Only branch support values over 0·85 are shown. Relevant branch support values are encircled. The scale bars indicate 0·1 expected changes per site. Name abbreviations are as in Fig. 6. Additional information for each protein is provided in Supplementary Table S1.

the Alphaproteobacteria and Acidobacteria, the most plausible explanation for the experimental evidence reported here seems to be an LGT event. In agreement with this line of reasoning, the incongruent placement of the Alphaproteobacteria in the LexA tree, clustering precisely with the Acidobacteria at the natural branching point of the latter, gives further credence to the hypothesis that an LGT event took place between the two clades. Furthermore, the fact that studies using RecA, UvrA, Ssb and UvrD proteins locate both *Acidobacterium* and the Alphaproteobacteria at their accepted branching points, in accordance with observations using well-established phylogenetic markers such as the 16s RNA, suggests that the transfer proposed here occurred from Acidobacteria, or an immediate ancestor presenting the GTTCN$_7$GTTC motif as its LexA box, to the Alphaproteobacteria. In accordance with this line of reasoning, an alignment of several LexA proteins that are representative of different bacterial groups (Fig. 1) reveals that the Alphaproteobacteria LexA protein possesses a 33 amino acid insertion downstream of the predicted third helix (α3) of its binding domain that is not seen in any other bacterial group. This gives further credence to the direction of the LGT deduced from the LexA tree, and also supports the hypothesis that this insert does not affect the DNA binding domain (DBD) (Knegtel *et al.*, 1995).

The data reported in this work also demonstrate that the *A. capsulatum recA* gene is DNA-damage inducible (Table 2), although the *A. capsulatum* LexA protein is unable to bind to the *recA* gene promoter (Fig. 5). A similar result has also been described for other bacterial species, such as *M. tuberculosis*, *Xylella fastidiosa* and *Myx. xanthus*, in which several DNA-repair-related genes (such as *uvrA* and *ssb*) are DNA-damage inducible, even though their expression is not directly under LexA protein control (Brooks *et al.*, 2001; Campoy *et al.*, 2002, 2003). However, to date, the mechanism controlling the LexA-independent DNA-damage-mediated induction of these genes is not known. Likewise, it is also not known if these bacterial species (*M. tuberculosis*, *X. fastidiosa* and *A. capsulatum*) possess a similar control mechanism to regulate the expression of DNA-damage-inducible genes that lack a LexA box in their promoters. Further work is necessary to understand how this *lexA*-independent gene induction takes place, as well as to determine its relevance in the bacterial response systems against DNA damage.

ACKNOWLEDGEMENTS

This work was funded by grant BFM2004-02768/BMC from the Ministerio de Educación y Ciencia (MEC) de España, and by the Consejo Superior de Investigaciones Científicas (CSIC). S. C. is a recipient of a postdoctoral contract from Instituto Nacional de Investigaciones y Tecnología Agraria y Alimentaria (INIA) – Institut de Recerca i Tecnologia Agroalimentàries (IRTA). We are deeply grateful to Joan Ruiz and Dr Pilar Cortés for their excellent technical assistance.

REFERENCES

Abella, M., Erill, I., Jara, M., Mazón, G., Campoy, S. & Barbé, J. (2004). Widespread distribution of a *lexA*-regulated DNA damage-inducible multiple gene cassette in the Proteobacteria phylum. *Mol Microbiol* **54**, 212–222.

Boshoff, H. I., Reed, M. B., Barry, C. D., 3rd & Mizrahi, V. (2003). DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. *Cell* **18**, 183–193.

Brooks, P. C., Movahedazadeh, F. & Davis, E. O. (2001). Identification of some DNA damage-inducible genes of *Mycobacterium tuberculosis*: apparent lack of correlation with LexA binding. *J Bacteriol* **183**, 4459–4467.

Bustin, S. A. (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR); trends and problems. *J Mol Endocrinol* **29**, 23–39.

Campoy, S., Mazón, G., Fernández de Henestrosa, A. R., Llagostera, M., Monteiro, P. B. & Barbé, J. (2002). A new regulatory DNA motif of the gamma subclass Proteobacteria: identification of the LexA protein binding site of the plant pathogen *Xylella fastidiosa*. *Microbiology* **148**, 3583–3597.

Campoy, S., Fontes, M., Padmanabhan, S., Cortes, P., Llagostera, M. & Barbe, J. (2003). LexA-independent DNA damage-mediated induction of gene expression in *Myxococcus xanthus*. *Mol Microbiol* **49**, 769–781.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540–552.

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.

Dubnau, D. & Lovett, C. M. (2002). Transformation and recombination. In *Bacillus Subtilis and its Closest Relatives: from Genes to Cells*, pp. 453–471. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.

Erill, I., Escribano, M., Campoy, S. & Barbé, J. (2003). *In silico* analysis reveals substantial variability in the gene contents of the Gamma Proteobacteria LexA-regulon. *Bioinformatics* **19**, 2225–2236.

Erill, I., Jara, M., Salvador, N., Escribano, M., Campoy, S. & Barbé, J. (2004). Differences in LexA regulon structure among Proteobacteria through in vivo assisted comparative genomics. *Nucleic Acids Res* **32**, 6617–6626.

Fernández de Henestrosa, A. R., Rivera, E., Tapias, A. & Barbé, J. (1998). Identification of the *Rhodobacter sphaeroides* SOS box. *Mol Microbiol* **28**, 991–1003.

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.

Fernández de Henestrosa, A. R., Cuñé, J., Erill, I., Magnuson, J. K. & Barbé, J. (2002). A green nonsulfur bacterium, *Dehalococcoides ethenogenes*, with the LexA binding sequence found in Gram-positive organisms. *J Bacteriol* **184**, 6073–6080.

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.

Griffiths, E. & Gupta, R. S. (2001). The use of signature sequences in different proteins to determine the relative branching order of bacterial divisions: evidence that *Fibrobacter* diverged at a similar time to *Chlamydia* and the *Cytophaga-Flavobacterium-Bacteroides* division. *Microbiology* **147**, 2611–2622.

Hall, T. A. (1999). BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.

Hanna, M. N., Ferguson, R. J., Li, Y.-H. & Cvitkovitch, D. G. (2001). *uvrA* is an acid-inducible gene involved in the adaptative response to low pH in *Streptococcus mutans*. *J Bacteriol* **183**, 5964–5973.

Hiraishi, A., Kishimoto, N., Kosako, Y., Wakao, N. & Tano, T. (1995). Phylogenetic position of the menaquinone-containing acidophilic chemo-organotroph *Acidobacterium capsulatum*. *FEMS Microbiol Lett* **132**, 91–94.

Kishimoto, N., Kosako, Y. & Tano, T. (1991). *Acidobacterium capsulatum* gen. nov., sp. nov. An acidophilic chemoorganotrophic bacterium containing menaquinone from acidic mineral environment. *Curr Microbiol* **22**, 1–7.

Knegtel, R. M. A., Fogh, R. H., Ottleben, G., Rüterjans, H., Dumoulin, P., Schnarr, M., Boelens, R. & Kaptein, R. (1995). A model for the LexA repressor DNA complex. *Proteins* **21**, 226–236.

Little, J. W. (1991). Mechanism of specific LexA cleavage: autodigestion and the role of RecA coprotease. *Biochimie* **73**, 411–422.

Little, J. W., Edminston, S., Pacelli, L. & Mount, D. (1980). Cleavage of the *Escherichia coli* LexA protein by the RecA protease. *Proc Natl Acad Sci U S A* **77**, 3225–3229.

Little, J. W., Mount, D. & Yanish-Perron, C. R. (1981). Purified LexA protein is a repressor of the *recA* and *lexA* genes. *Proc Natl Acad Sci U S A* **78**, 4199–4203.

Ludwig, W., Bauer, S. H., Bauer, M. & 7 other authors (1997). Detection and *in situ* identification of representatives of a widely distributed new bacterial phylum. *FEMS Microbiol Lett* **153**, 181–190.

Mazón, G., Lucena, J. M., Campoy, S., Fernández de Henestrosa, A. R., Candau, P. & Barbé, J. (2004a). LexA-binding sequences in Gram-positive and cyanobacteria are closely related. *Mol Gen Genomics* **271**, 40–49.

Mazón, G., Erill, I., Campoy, S., Cortés, P., Forano, E. & Barbé, J. (2004b). Reconstruction of the evolutionary history of the LexA-binding sequence. *Microbiology* **150**, 3783–3795.

Notredame, C., Higgins, D. & Heringa, J. (2000). T-Coffee: a novel method for multiple sequence alignments. *J Mol Biol* **302**, 205–217.

Page, R. D. M. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *CABIOS* **12**, 357–358.

Quaiser, A., Ochsenreiter, C. L., Schuster, S. C., Treusch, A. H., Eck, J. & Schleper, C. (2003). Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Mol Microbiol* **50**, 563–575.

Raja, N., Goodson, M., Chui, W. C., Smith, D. G. & Rowbury, R. J. (1991). Habituation to acid in *Escherichia coli*: conditions for habituation and its effects on plasmid transfer. *J Appl Bacteriol* **70**, 59–65.

Ramesar, R. S., Abratt, V., Woods, D. R. & Rawlings, D. E. (1989). Nucleotide sequence and expression of a cloned *Thiobacillus ferrooxidans recA* gene in *Escherichia coli*. *Gene* **78**, 1–8.

Ronquist, F. & Huelsenbeck, J. P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1992). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sanger, F., Nicklen, S. & Coulson, S. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463–5467.

Sassanfar, M. & Roberts, J. W. (1990). Nature of SOS-inducing signal in *Escherichia coli*. The involvement of DNA replication. *J Mol Biol* **212**, 79–96.

Tapias, A. & Barbé, J. (1999). Regulation of divergent transcription from the *uvrA-ssb* promoters in *Sinorhizobium meliloti*. *Mol Gen Genet* **262**, 121–130.

Tapias, A., Fernández, S., Alonso, J. C. & Barbé, J. (2002). *Rhodobacter sphaeroides* LexA has dual activity: optimising and repressing *recA* gene transcription. *Nucleic Acids Res* **30**, 1539–1546.

Thompson, S. A. & Blaser, M. J. (1995). Isolation of the *Helicobacter pylori recA* gene and involvement of the *recA* region in resistance to low pH. *Infect Immun* **63**, 2185–2193.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Tomb, J. F., White, O., Kerlavage, A. R. & 39 other authors (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539–547.

Walker, G. C. (1984). Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. *Microbiol Rev* **48**, 60–93.

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: redefinition of the consensus sequence. *J Bacteriol* **180**, 2201–2211.