The *Verrucomicrobia* LexA-Binding Motif: Insights into the Evolutionary Dynamics of the SOS Response

Ivan Erill[1], Susana Campoy[2], Sefa Kılıç[1] and Jordi Barbé[2]*

[1] Erill Lab, Department of Biological Sciences, University of Maryland Baltimore County, Baltimore, MD, USA, [2] Unitat de Microbiologia, Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Barcelona, Spain

The SOS response is the primary bacterial mechanism to address DNA damage, coordinating multiple cellular processes that include DNA repair, cell division, and translesion synthesis. In contrast to other regulatory systems, the composition of the SOS genetic network and the binding motif of its transcriptional repressor, LexA, have been shown to vary greatly across bacterial clades, making it an ideal system to study the co-evolution of transcription factors and their regulons. Leveraging comparative genomics approaches and prior knowledge on the core SOS regulon, here we define the binding motif of the *Verrucomicrobia*, a recently described phylum of emerging interest due to its association with eukaryotic hosts. Site directed mutagenesis of the *Verrucomicrobium spinosum recA* promoter confirms that LexA binds a 14 bp palindromic motif with consensus sequence TGTTC-N4-GAACA. Computational analyses suggest that recognition of this novel motif is determined primarily by changes in base-contacting residues of the third alpha helix of the LexA helix-turn-helix DNA binding motif. In conjunction with comparative genomics analysis of the LexA regulon in the *Verrucomicrobia* phylum, electrophoretic shift assays reveal that LexA binds to operators in the promoter region of DNA repair genes and a mutagenesis cassette in this organism, and identify previously unreported components of the SOS response. The identification of tandem LexA-binding sites generating instances of other LexA-binding motifs in the *lexA* gene promoter of *Verrucomicrobia* species leads us to postulate a novel mechanism for LexA-binding motif evolution. This model, based on gene duplication, successfully addresses outstanding questions in the intricate co-evolution of the LexA protein, its binding motif and the regulatory network it controls.

Keywords: SOS response, LexA regulon, DNA repair, comparative genomics, binding motif, translesion synthesis, uracil-DNA glycosylase, regulatory network evolution

INTRODUCTION

The SOS response is the primary mechanism for coordinating the response to DNA damage in *Bacteria* (Erill et al., 2007). First reported in *Escherichia coli*, (Little and Mount, 1982), the SOS response has been documented in a broad range of bacterial species (Erill et al., 2007). In *E. coli* and *Bacillus subtilis*, the SOS response has been shown to regulate between 30 and 40 genes involved in DNA repair, translesion synthesis, and cell-division arrest (Fernandez De Henestrosa et al., 2000; Walker et al., 2000; Au et al., 2005). This regulatory network is governed by the transcriptional

repressor LexA, which in *E. coli* binds as a homodimer to specific sites upstream of regulated operons and blocks transcription initiation (Thliveris et al., 1991; Walker et al., 2000). Upon DNA damage, the RecA protein binds single-stranded DNA (ssDNA) fragments originating at stalled replication forks, and forms active nucleoprotein filaments capable of promoting self-cleavage of the LexA repressor (Sassanfar and Roberts, 1990). Self-cleavage of the LexA dimer leads to de-repression of target operons, which typically include the *lexA* and *recA* genes (Little, 1991), and full induction of the system (Walker et al., 2000). In recent years, the SOS response has attracted increasing interest due to its active involvement in the regulation of mobile genetic elements, such as integrative and conjugative elements (Beaber et al., 2004), pathogenicity islands (Ubeda et al., 2007), and integron integrases (Guerin et al., 2009), as well as its induction by different types of antibiotics (Beaber et al., 2004; Ubeda et al., 2005; Maiques et al., 2006).

Beyond its clinical interest, the SOS response also constitutes a unique model for the study of the evolution of transcriptional regulatory networks. In contrast with many other transcriptional regulators, the LexA repressor displays remarkably different binding motifs across multiple phyla, changing both the specificity of the dyad region recognized by each LexA monomer as well as the dyad space (Erill et al., 2007). Reported LexA-binding motifs range from short inverted repeats (GAAC-N4-GTTC) in the *Firmicutes* and *Actinobacteria* (Davis et al., 2002; Au et al., 2005), to larger palindromic motifs (CTGT-N8-ACAG) in the *Gammaproteobacteria* (Fernandez De Henestrosa et al., 2000; Erill et al., 2003) and even direct repeat motifs (GTTC-N7-GTTC) in the *Alphaproteobacteria* (Fernandez de Henestrosa et al., 1998; Erill et al., 2004). This variability in LexA-binding motifs is matched by extreme plasticity in the size and composition of the SOS regulatory network, which can regulate from 3 to 40 genes (Fernandez De Henestrosa et al., 2000; Au et al., 2005; Campoy et al., 2005) and has been shown to broadly comprise a minimal shared SOS regulon core consisting of *lexA*, *recA*, and a mutagenesis gene cassette (*imuA-imuB-dnaE2*; Erill et al., 2006).

Named after *Verrucomicrobium spinosum*, the *Verrucomicrobia* are a recently established bacterial phylum characterized by species with distinct wart-like morphology (Garrity and Holt, 2001) and divided in three main classes (Opitutae, Spartobacteria, and *Verrucomicrobiae*; Bergmann et al., 2011). *Verrucomicrobia* possess several unusual features, like the presence of a eukaryotic-like tubulin (Schlieper et al., 2005), but interest in this phylum has grown in recent years due mainly to metagenomics analyses revealing the association of *Verrucomicrobia* with several eukaryotic hosts (Sait et al., 2011), their prominence in many soil communities (Bergmann et al., 2011) and a significant role in the adaptability of the human gut microbiome (Dubourg et al., 2013; Liou et al., 2013). *Verrucomicrobia* are clustered with the *Planctomycetes* and the *Chlamydiae* in the *Planctomycetes-Verrucomicrobia-Chlamydiae* (PVC) super-phylum, a large and diverse phylogenetic clade of clinical and biotechnological interest in which the SOS response has not been documented (Gupta et al., 2012). The genome of the representative *Verrucomicrobia* species, *V. spinosum*, reveals

the presence of orthologs for the three core constituents of the SOS response (*lexA*, *recA*, and the *imuA-imuB-dnaE2* operon), suggesting that the *Verrucomicrobia* have a functional SOS response. Here we combine *in silico* and *in vitro* approaches to characterize the LexA-binding motif of *Verrucomicrobia* and analyze the SOS regulatory network of this bacterial phylum. Our results illustrate the extraordinary plasticity of this transcriptional regulatory network and provide novel insights into the molecular mechanisms driving its evolution.

MATERIALS AND METHODS

Functional and Taxonomical Assignments

Orthology assignment of putative function for all reported genes was performed using the HHPRED web service (HHpred, RRID:SCR_010276) with default options and an *e*-value threshold of 10^{-30} on the TIGRFAM (JCVI TIGRFAMS, RRID:SCR_005493), PFAM (Pfam, RRID:SCR_004726), and COG (COG, RRID:SCR_007139) databases (Tatusov et al., 2000; Söding et al., 2005; Haft et al., 2013; Finn et al., 2016). eggNOG identifiers, categories, and descriptions were retrieved from the eggNOG database (eggNOG, RRID:SCR_002456) using HMMER (Hmmer, RRID:SCR_005305; Eddy, 2011; Powell et al., 2014). Taxonomical identification numbers (TaxID) for all species analyzed in the manuscript were retrieved from the NCBI Taxonomy server (NCBI Taxonomy, RRID:SCR_003256; Federhen, 2012).

Ortholog Detection and Motif Discovery

Protein sequences for *V. spinosum* DSM 4136 [TaxID: 240016] LexA [WP_009959117], RecA [WP_009965965], and ImuA [WP_009959308] were obtained from the NCBI RefSeq database [RefSeq, RRID:SCR_003496; Pruitt et al., 2012]. The promoter regions ([−250, 0] with respect to the translational start site) for orthologs of these sequences in *Verrucomicrobia* genome and metagenome assemblies were downloaded from the Integrated Microbial Genomes (IMG) resource [IMG, RRID:SCR_007733; Markowitz et al., 2014], after identifying orthologs through the IMG BLAST service using a 10^{-20} *e*-value threshold and the *V. spinosum* sequences as queries (Altschul et al., 1997). Downloaded promoter sequences were used as input for the MEME motif discovery tool on the MEME Suite server [MEME Suite—Motif-based sequence analysis tools, RRID:SCR_001783], requesting palindromic motifs between 8 and 20 bp long and otherwise default options (Bailey et al., 2006). Sequence logos were generated using the Weblogo sequence service (WEBLOGO, RRID:SCR_010236; Crooks et al., 2004).

Analysis Of α3 Helix Motifs

A sequence model for the *Verrucomicrobia* α3 helix of the LexA N-terminal helix-turn-helix motif was obtained through multiple sequence alignment of available LexA protein sequences for this phylum, using the information in P0A7C2 (UniProtKB, RRID:SCR_004426) on the *E. coli* LexA crystal structure (Zhang et al., 2010) to define penalty masks for CLUSTALW profile alignment mode (Thompson et al., 1994). The *Verrucomicrobia* LexA α3 helix motif was compared to previously compiled LexA

α3 helix motifs for different phyla (Sanchez-Alberola et al., 2015) using the TomTom service of the MEME suite (Gupta et al., 2007). Amino acid property plots and differential analysis of the *Verrucomicrobia* LexA α3 helix motif with respect to the *Betaproteobacteria*, *Actinobacteria,* and *Firmicutes* α3 helix motifs were generated with the iceLogo web service (iceLogo, RRID:SCR_012137; Colaert et al., 2009).

LexA-Binding Motif Search and Comparative Genomics Analysis

Experimentally validated LexA-binding motifs were downloaded from the CollecTF database (CollecTF, RRID:SCR_014405; Kiliç et al., 2014). Whole genome shotgun assemblies for *Verrucomicrobia* species with total sequence lengths larger than or equal to the smallest complete *Verrucomicrobia* genome (*Methylacidiphilum infernorum* V4, TaxID: 481448; 2,287,145 bp) were obtained from the NCBI RefSeq database (RefSeq, RRID:SCR_003496). LexA-binding motif searches on individual genomes were performed using xFITOM (xFITOM, SCR_014445) with the sequence information content (R_i) scoring method and default parameters (Schneider, 1997; Bhargava and Erill, 2010). Comparative genomics analyses were performed with CGB, a collection of Python scripts implementing a computational pipeline for comparative genomics of regulatory transcriptional networks in bacterial genomes. The pipeline is based on previous work (Sanchez-Alberola et al., 2015) and is available under a GPL license on GitHub (http://www.github.com/erilllab/cgb). Given a set of genome assemblies, a transcription factor (TF) and its known binding motif, the pipeline first searches for binding motif instances in the promoter region of all genes (-250, $+50$ of TLS). Genes in directons with intergenic distance below the mean intergenic distance of each genome are considered to form operons and TF-binding sites identified in the lead operon gene are assigned accordingly to all operon members. The presence of genes with high-scoring TF-binding sites within predicted operons is used to revise operon predictions. Orthologs across all analyzed species are detected as best reciprocal BLAST hits using a 10^{-20} *e*-value threshold. The pipeline summarizes analysis results using a heatmap with species clustered using a distance-based TF tree and a color scheme indicating the presence/absence of orthologs and the score of detected TF-binding sites in the corresponding operon.

Bacterial Strains and Culture Conditions

E. coli (DH5α and BL21; Thermo Fisher Scientific, RRID:SCR_013270) and *V. spinosum* DSM 4136 strains were grown at 37°C in LB (Green and Sambrook, 2012) and at 30°C in M13 media [DSMZ 607; German Collection of Microorganisms and Cell Cultures, RRID:SCR_001711], respectively. Antibiotics were added to the cultures at reported concentrations (Green and Sambrook, 2012).

Oligonucleotides and DNA Techniques

Plasmid isolation, restriction digestion, DNA ligation, transformation, DNA extraction, and PCR were carried out using standard protocols, as described elsewhere (Green and

Sambrook, 2012). Restriction enzymes, T4 DNA ligase, DNA polymerase, and the DIG-DNA labeling and detection kit were from Roche (Roche NimbleGen, RRID:SCR_008571). The oligonucleotides used for this work are listed in Supplementary Material 1 and were purchased from Invitrogen (Molecular Probes, RRID:SCR_013318). Mutants of the *V. spinosum recA* promoter (VSP_RS32310) were obtained using oligonucleotides carrying designed substitutions (Supplementary Material 1). The DNA sequence of generated fragments was verified by sequencing (Macrogen, RRID:SCR_014454).

Protein Purification and Electrophoresis Mobility Shift Assays

V. spinosum DSM 4136 DNA was extracted from phosphate buffered (50 mM) saline (pH 8.0)-washed pellets containing cells using the easy-DNA™ DNA isolation kit (Molecular Probes, RRID:SCR_013318). The *V. spinosum* DSM 4136 was amplified using suitable primers (Supplementary Material 1) and cloned into a pET15b vector (Millipore, RRID:SCR_008983). The *O. terrae* PB90-1 lexA was obtained by chemical synthesis (GeneArt; Thermo Fisher Scientific, RRID:SCR_013270) and cloned into a pET15b vector. Overexpression and purification of the *V. spinosum* DSM 4136 and *O. terrae* PB90-1 and *B. subtilis* LexA proteins was performed as described previously for other LexA proteins (Cambray et al., 2011; Cornish et al., 2014). DNA probes for electro-mobility shift assays (EMSA) were generated using two complementary synthetic oligonucleotides centered on the target LexA-binding sites (Supplementary Material 1). The dsDNA synthetic fragments were ligated into pGEMT vector (Roche NimbleGen, RRID:SCR_008571) and transformed into *E. coli* DH5α (Thermo Fisher Scientific, RRID:SCR_013270). In all cases the plasmids were confirmed by sequencing and DNA probes were obtained by PCR using M13 forward and reverse digoxigenin-labeled oligos (Supplementary Material 1). EMSAs were performed as described previously (Sanchez-Alberola et al., 2012), using 20 ng of each digoxigenin-marked DNA probe in the binding mixture and adding the corresponding LexA protein (from 80 to 400 nM). Samples were loaded onto 6% non-denaturing Tris-glycine polyacrylamide gels and digoxigenin-labeled DNA-protein complexes were detected using the manufacturer's protocol (Roche NimbleGen, RRID:SCR_008571).

RESULTS

LexA Targets a Novel LexA-Binding Motif in the *Verrucomicrobia*

The presence of core SOS response operons (*lexA* [VSP_RS04780], *recA* [VSP_RS32310], and *imuA-imuB-dnaE2* [VSP_RS05590-VSP_RS05595-VSP_RS05600]) in the genome of the representative *Verrucomicrobia* species *V. spinosum* DSM 4136 indicates that this phylum might possess a functional LexA regulatory network. However, computational searches using known LexA-binding motifs did not yield putative LexA-binding sites upstream of any SOS related genes in *V. spinosum*. Taking advantage of the availability of multiple

genome and metagenome assemblies for the *Verrucomicrobia* phylum, we compiled 116 promoter sequences from 59 different assemblies corresponding to orthologs of the *V. spinosum* DSM 4136 *lexA*, *recA,* and *imuA* genes through the JGI-IMG service (Supplementary Material 2). We then used MEME to identify overrepresented motifs in these sequences. The most significant motif identified by MEME (**Figure 1A**) is a 14 bp palindromic motif with consensus sequence (TGTTC-N4-GAACA). This motif was identified in the promoter region of 27 *lexA* genes, 25 *recA* genes, and 3 *imuA* genes, corresponding to 36 different genome and metagenome assemblies and spanning all three major groups of *Verrucomicrobia* (Supplementary Material 3). A computational search also identified instances of this motif in the promoter sequences of the *V. spinosum* DSM 4136 *lexA*, *recA,* and *imuA-imuB-dnaE2* operons (**Figure 1B**). The TGTTC-N4-GAACA motif is reminiscent of the LexA-binding motif (GAAC-N4-GTTC) previously reported in the *Firmicutes*, *Actinobacteria,* and Gallionellales (Davis et al., 2002; Au et al., 2005; Sanchez-Alberola et al., 2015). Together with its structural similarity to previously reported LexA-binding motifs, the presence of this motif in the promoter region of multiple orthologs for three core components of the SOS response

strongly suggested that the identified TGTTC-N4-GAACA motif is the LexA-binding motif of the *Verrucomicrobia*.

To validate that the palindromic motif identified *in silico* was the LexA-binding motif of *V. spinosum*, we purified the *V. spinosum* DSM 4136 LexA protein [WP_009959117] and performed electro-mobility shift assays (EMSA) with wild-type and mutant versions of the *V. spinosum* DSM 4136 *recA* promoter containing single-nucleotide substitutions at each position of the predicted LexA-binding motif. The results of this site-directed mutagenesis analysis (**Figure 1C**) are in broad agreement with the motif predicted *in silico*, confirming that *V. spinosum* LexA targets a spaced dyad motif with consensus sequence TGTTC-N4-GAACA. Single-nucleotide mutations to the bases of the inverted repeat regions (TGTTC and GAACA) of the *V. spinosum* LexA-binding motif systematically abolish LexA binding in the *recA* promoter context, indicating that these conserved elements likely correspond to the monomer binding site and are therefore essential for LexA binding activity (Groban et al., 2005). In contrast, the 4 bp spacer region and 3 bp flanking regions tolerate single-nucleotide mutations, suggesting that they are predominantly involved in indirect readout and DNA bending (Zhang et al., 2010).
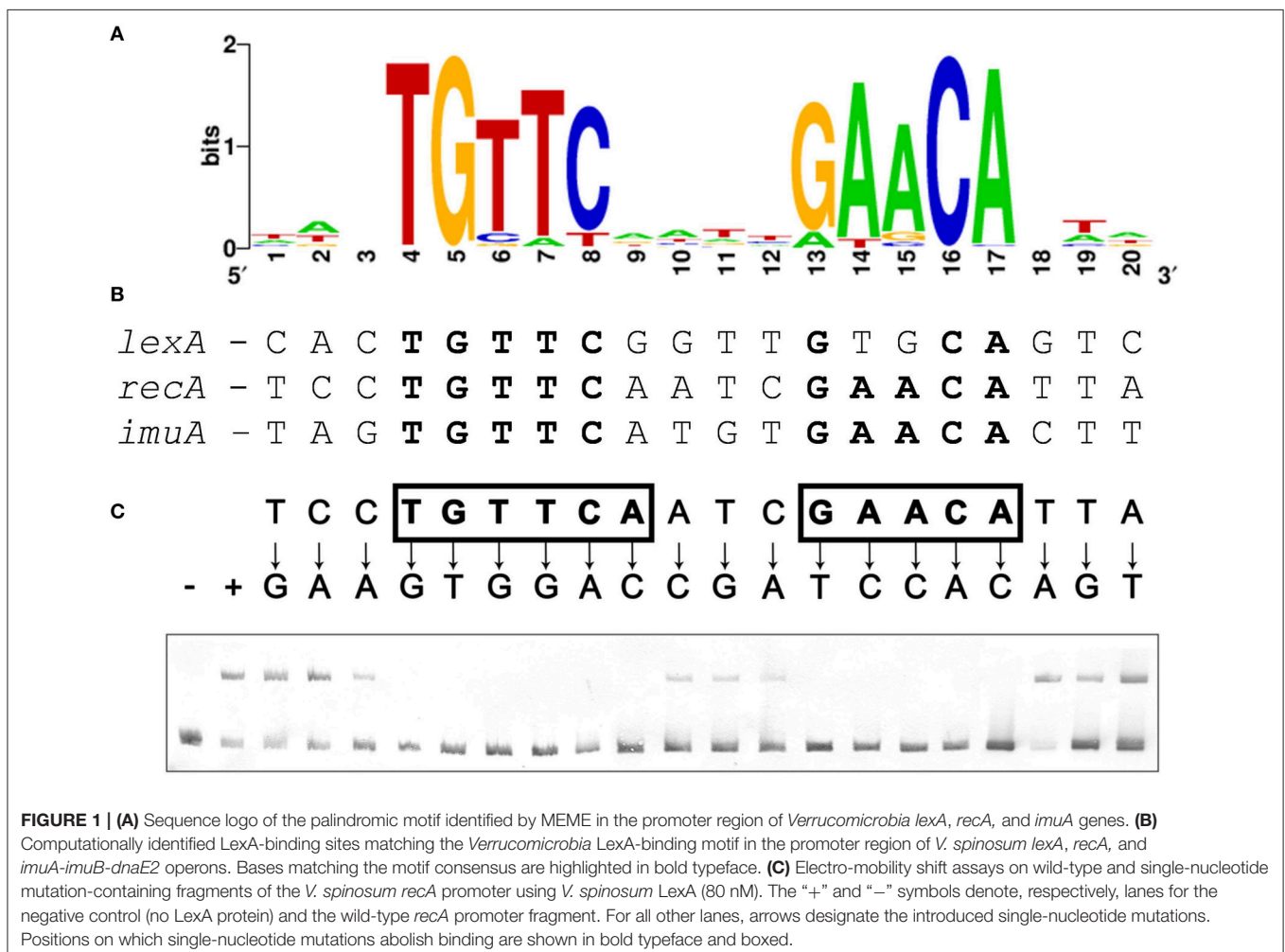
FIGURE 1 | (A) Sequence logo of the palindromic motif identified by MEME in the promoter region of *Verrucomicrobia lexA*, *recA,* and *imuA* genes. **(B)** Computationally identified LexA-binding sites matching the *Verrucomicrobia* LexA-binding motif in the promoter region of *V. spinosum lexA*, *recA,* and *imuA-imuB-dnaE2* operons. Bases matching the motif consensus are highlighted in bold typeface. **(C)** Electro-mobility shift assays on wild-type and single-nucleotide mutation-containing fragments of the *V. spinosum recA* promoter using *V. spinosum* LexA (80 nM). The "+" and "−" symbols denote, respectively, lanes for the negative control (no LexA protein) and the wild-type *recA* promoter fragment. For all other lanes, arrows designate the introduced single-nucleotide mutations. Positions on which single-nucleotide mutations abolish binding are shown in bold typeface and boxed.

Previous work has established that the α3 helix of the N-terminal helix-turn-helix motif is responsible for the majority of the specific contacts with monomer binding sites of LexA-binding motifs (Oertel-Buchheit et al., 1990; Ottleben et al., 1991; Thliveris and Mount, 1992; Groban et al., 2005; Zhang et al., 2010). Comparison of the α3 helix sequence in *Verrucomicrobia* LexA proteins with previously reported LexA α3 helix motifs (Sanchez-Alberola et al., 2015) shows that the *Verrucomicrobia* LexA α3 helix is most closely related to those of the *Betaproteobacteria*, *Firmicutes*, and *Actinobacteria*. As shown in **Figure 2**, the majority of the changes observed in the α3 helix of *Verrucomicrobia* localize to the N-terminal part of the helix, affecting residues that change sequence specificity through direct readout, but that are not essential for DNA bending and structural motif recognition (Oertel-Buchheit et al., 1990; Thliveris et al., 1991; Thliveris and Mount, 1992; Groban et al., 2005; Zhang et al., 2010). Furthermore, the overall distribution of hydrogen donors and hydrophobic residues is preserved across the entire α3 helix

(Supplementary Material 4). These observations suggest that the structural similarities between *Firmicutes*, *Actinobacteria*, *Betaproteobacteria,* and *Verrucomicrobia* LexA-binding motifs are the result of an evolutionary process in the LexA DNA-binding motif that has modified the specific readout of monomer sites without altering the recognition of the overall motif structure.

The *Verrucomicrobia* LexA Protein Targets Tandem Binding Sites in *LexA* Promoters

Close inspection of the *V. spinosum lexA* promoter reveals a poorly conserved LexA-binding site immediately downstream (1 bp) of the putative LexA-binding site identified *in silico* (**Figure 3**). To confirm that both these putative motif instances are involved in LexA binding, we performed EMSA with purified *V. spinosum* LexA protein on the *lexA* promoter. The results shown in **Figure 4A** revealed the distinct formation of two retardation bands on the *lexA* promoter at low protein concentrations, corresponding to LexA binding at either one

FIGURE 2 | (A) Sequence logo of the *Verrucomicrobia* LexA α3 helix motif. **(B)** IceLogo depicting the difference in amino acid frequency at each position of the LexA α3 helix between the multiple sequence alignment of *Verrucomicrobia* LexA protein sequences and a reference set of *Betaproteobacteria*, *Actinobacteria,* and *Firmicutes* LexA α3 helix sequences. The upper part of the plot shows residues overrepresented in the *Verrucomicrobia* LexA α3 helix; the bottom part shows residues overrepresented in the reference set. Only differences with significant *z*-score under a confidence interval of 0.01 are shown.

```
V. spinosum DSM 4136          GACTGCACAACCGAACAGTGTTCACATCTAACCATGCTTACGGA
V. sp. BvORR106               GACTGCACAATCGAACAGTGTTCACATCTAACCATGCTTACGGA
V. sp. BvORR034               GACTGCACAATCGAACAGTGTTCACATCTAACCATGCTTACAGA
R. marina DSM 17716           CTATGTTCTTAAGAACAGTGTTCAAATTATGAGTGATGCATTAA
R. squalenifaciens DSM 18772  TTGTGTTCTTAAGAACAGTGTTCAATTAGGTCATGGCAAAAGGA
V. SCGC AAA027-I19            TTTTGTTCTGTCGAACACTGTTCATCAGATGAGCTCAAACCAAC
MEint.metabat.12173           TTTTGTTCTGTCGAACACTGTTCATCAGATGAGCTCAAACCAAC
MEint.metabat.12612           ATTTGTTCTCAAGAACACTGTTCAACGAACCATGGCACACAACC
TBhypo.metabat.2519           TAGTGTTCAAAAGAACACTGTTCATATGAGTAAGCAATTAACGG
TBepi.metabat.1800            TAGTGTTCAAAAGAACACTGTTCATATGAGTAAGCAATTAACGG
MEint.metabat.3880            CTTTGTTCCCTCGAACACTGTTCAAAATCATGCACACTAACCTC
Haloferula BvORR071           GTTTGTTCCTAAGAACACTGTTCACTAAGCCATGCACGAAGGGT
V. EBPR_Bin_208               GAATGTTCTCTCAAACAGTGTTCAGAATGAAGTTAACCTCACGC
V. EBPR_Bin_287               TGATGTTCTGTGGAACAGTGTTCATAATGAAATTAACCGCCCGC
V. IMCC26134                  ATGTGTTCACTTGAACACGGTTCACAAGAACACACCGCCATGCA
Opitutae-129                  AGGTGTTCATTTGAACCGTGTTCAATTGAACCCTATGTTAACCG

O. terrae PB90-1              GGCTGTTCGCTTGAACATGTGTTCAAGAGAACACATCCGCCATG
```

FIGURE 3 | Multiple sequence alignment of *Verrucomicrobia lexA* promoter sequences displaying the tandem LexA-binding site arrangement observed in *V. spinosum*. The primary LexA-binding site is highlighted in yellow, and the secondary one in green. Bases matching the LexA-binding motif consensus are in bold typeface. Predicted translation start sites are shown in red and underlined.
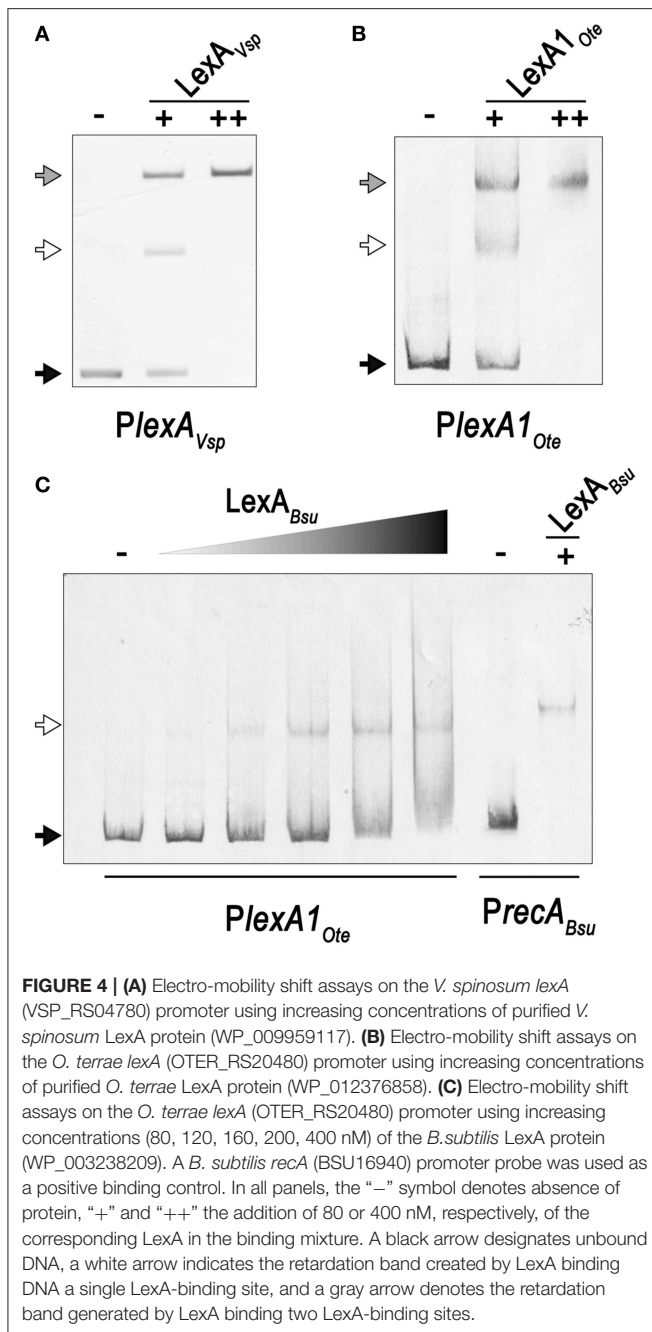
or the two LexA-binding sites identified in the *lexA* promoter. Further, increasing protein concentration resulted in a single retardation band corresponding to LexA recognizing both LexA-binding sites. Taken together, these results indicate that the two identified LexA-binding sites in the promoter region of the *V. spinosum lexA* gene are bound cooperatively by LexA. A systematic analysis of the promoter regions of *V. spinosum lexA* gene orthologs in the *Verrucomicrobia* revealed that more than half of the *lexA* ortholog promoters with predicted LexA-binding sites display similar tandem site configurations (Supplementary Material 5). Most of these tandem arrangements involve a conserved TGTTC-N4-GAACA motif instance followed by a degenerate site in which only the first TGTTC element is conserved, but a tandem site arrangement with both conserved sites can be observed in at least two species (**Figure 3**).

In *Opitutus terrae* PB90-1, there are two fully conserved *Verrucomicrobia* LexA-binding motifs in the promoter region of a putative *lexA-imuA-imuB-dnaE2* operon [OTER_RS20480-OTER_RS20475-OTER_RS20470-OTER_RS20465] separated by 2 bp. This arrangement generates an instance of the canonical GAAC-N4-GTTC LexA-binding motif of *Firmicutes* and *Actinobacteria*. Using purified *O. terrae* and *B. subtilis* LexA proteins, we performed EMSA to validate the functionality of this tandem arrangement in *O. terrae* (**Figure 4**). EMSA with *O. terrae* LexA [WP_012376858] reveals two retardation bands at low protein concentration, confirming that this protein also binds to both *Verrucomicrobia* LexA target sites in the *lexA-imuA-imuB-dnaE2* promoter (**Figure 4B**). Mobility assays with incremental concentrations of *B. subtilis* LexA [WP_003238209] show that *B. subtilis* LexA binds a unique element in the *O.*

terrae lexA-imuA-imuB-dnaE2 promoter, yielding a single retardation band similar to the one observed on the *B. subtilis recA* [BSU16940] promoter (**Figure 4C**). These results suggest that *B. subtilis* LexA binds the *Firmicutes*-like LexA-binding motif instance generated by the tandem arrangement of *Verrucomicrobia* LexA-binding sites.

The Core *Verrucomicrobia* LexA Regulon Comprises Three Operons Involved in DNA Repair and Mutagenesis

Having established the LexA-binding motif of the *Verrucomicrobia*, we performed a comparative genomics analysis of the LexA regulon in this phylum. We compiled 15 whole-genome shotgun assemblies for members of all the major classes of *Verrucomicrobia* (*Opitutae, Spartobacteria,* and *Verrucomicrobiae*) presenting a *V. spinosum* LexA homolog and searched for putative LexA-binding sites in the promoter region [−250, +50] of predicted operons. The results of the comparative genomics analysis (**Figure 5**; Supplementary Materials 6, 7) reveal a core LexA regulon present in all classes of the *Verrucromicrobia* phylum and composed of three operons: *lexA, splB,* and *imuA-imuB-dnaE2*. The *lexA* gene displays high-scoring sites in all the analyzed species, except for *Verrucomicrobium* sp. 3C (TaxID: 1134055), *Verrucomicrobia bacterium* LP2A (TaxID: 478741), and *Pedosphaera parvula* Ellin514 (TaxID: 320771). The LexA proteins of these species display significant changes to the α3 helix of the LexA DNA-binding domain, suggesting that they may target a divergent LexA-binding motif. Consistent with this

FIGURE 4 | (A) Electro-mobility shift assays on the *V. spinosum lexA* (VSP_RS04780) promoter using increasing concentrations of purified *V. spinosum* LexA protein (WP_009959117). **(B)** Electro-mobility shift assays on the *O. terrae lexA* (OTER_RS20480) promoter using increasing concentrations of purified *O. terrae* LexA protein (WP_012376858). **(C)** Electro-mobility shift assays on the *O. terrae lexA* (OTER_RS20480) promoter using increasing concentrations (80, 120, 160, 200, 400 nM) of the *B.subtilis* LexA protein (WP_003238209). A *B. subtilis recA* (BSU16940) promoter probe was used as a positive binding control. In all panels, the "−" symbol denotes absence of protein, "+" and "++" the addition of 80 or 400 nM, respectively, of the corresponding LexA in the binding mixture. A black arrow designates unbound DNA, a white arrow indicates the retardation band created by LexA binding DNA a single LexA-binding site, and a gray arrow denotes the retardation band generated by LexA binding two LexA-binding sites.
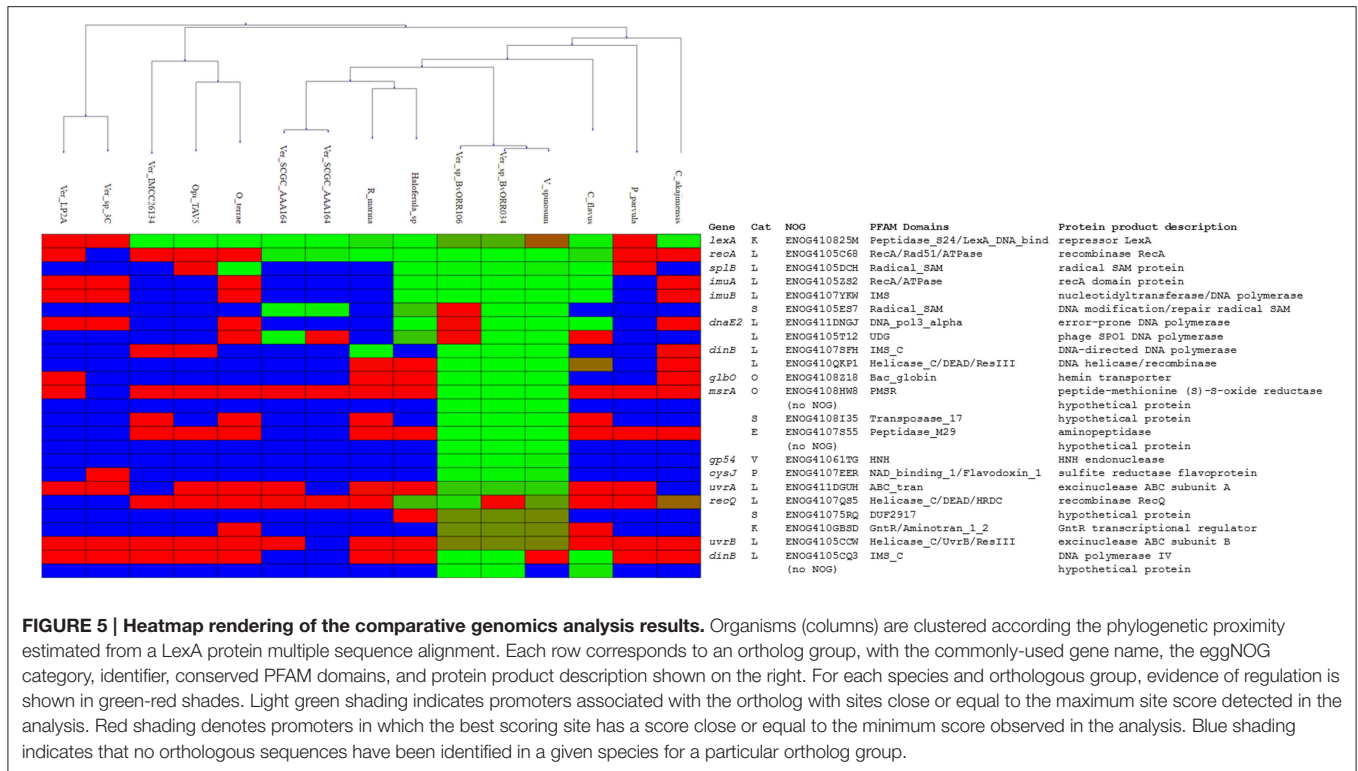
result, the genomes of these organisms do not reveal any instance of the *Verrucomicrobia* LexA-binding motif in the promoter regions of previously documented SOS genes (Erill et al., 2007). The promoter region of the *splB* gene shows evidence of LexA regulation in several *Verrucomicrobiae*, one *Opitutaceae* (*O. terrae*) and the only available assembly of a *Spartobacteria* species (*Chthoniobacter flavus* Ellin428; TaxID: 497964). The product of the *splB* gene contains a radical SAM domain (PFAM04055) and has homology to COG1533 (ENOG4105DCH), classified as a DNA repair photolyase. Members of this orthologous group have been reported to be regulated by LexA in the *Actinobacteria*,

the *Gammaproteobacteria*, the *Betaproteobacteria,* and the *Alphaproteobacteria* (Davis et al., 2002; Cirz et al., 2006; Sanchez-Alberola et al., 2012, 2015; Ulrich et al., 2013), suggesting that it may be a previously unrecognized core component of the SOS response. Lastly, the promoter region of the *imuA-imuB-dnaE2* operon presents *Verrucomicrobia* LexA-binding motif instances in *C. flavus* and the same *Verrucomicrobiae* species as *splB*. As noted above, *O. terrae* presents a putative *lexA-imuA-imuB-dnaE2* with verified *O. terrae* LexA binding in its promoter region (**Figure 4**). Even though the intergenic distance between *lexA* and *imuA* is larger than the genomic average for this species (264 bp), the prevalence of *lexA-imuA-imuB-dnaE2* arrangements across the *Bacteria* domain suggests that this direction constitutes a functional operon in *O. terrae* (Erill et al., 2006).

The *Verrucomicrobia* LexA Regulon is Highly Variable and Incorporates Novel Functions

The results of the comparative genomics analysis reveal remarkable variation in the size and composition of the inferred LexA regulon. In the *Verrucomicrobiae*, the predicted regulon ranges from one operon (*Verrucomicrobia bacterium* IMCC26134; TaxID: 1637999) to over 14 (*V. spinosum*), with several species displaying intermediate sizes [2 operons in *Rubritalea marina* DSM 17716 (TaxID: 1123070) or 5 in *Haloferula* sp. BvORR071 (TaxID: 1396141)]. The only available *Spartobacteria* representative (*C. flavus*) shows a moderate regulon size (5 operons). In contrast, the LexA regulon appears to have shrunk noticeably in the *Opitutaceae*, where it encompasses at the most two operons. Two members of this family [*O. terrae* and *Opitutaceae bacterium* TAV5 (TaxID: 794903)] present a duplication of the *lexA* gene. The products of the two *Opitutaceae* TAV5 *lexA* genes (OPIT5_RS22040 and OPIT5_RS25725) present 91% identity and their promoter regions contain almost identical LexA-binding sites. The *lexA* genes in *O. terrae* (OTER_RS20480 and OTER_RS11645) have diverged substantially (42% protein sequence identity) and only the promoter of the *lexA1* gene (OTER_RS20480) presents *Verrucomicrobia* LexA-binding motif instances, following the tandem arrangement discussed above (**Figure 3**).

The overall composition of the inferred *Verrucomicrobia* LexA regulon is in broad agreement with experimental and computational descriptions of the SOS regulatory network in other phyla (Erill et al., 2007; Sanchez-Alberola et al., 2012, 2015). Beyond the core regulon described above (*lexA*, *splB*, *imuA-imuB-dnaE2*), the *Verrucomicrobia* LexA regulon encompasses genes coding for the recombination protein RecA (COG0468; ENOG4105C68), the excinuclease ABC subunits A (COG0178; ENOG411DGUH) and B (COG0556; ENOG4105CCW), two DNA helicase RecQ homologs (COG0514; ENOG4107QS5 and ENOG410QKP1) and two homologs of the error-prone DNA polymerase IV (COG0389; ENOG4105CCW and ENOG4105CQ3). In addition to these previously established SOS genes, the *Verrucomicrobia* LexA regulon shows evidence of regulation for an operon encoding proteins matching the

FIGURE 5 | Heatmap rendering of the comparative genomics analysis results. Organisms (columns) are clustered according the phylogenetic proximity estimated from a LexA protein multiple sequence alignment. Each row corresponds to an ortholog group, with the commonly-used gene name, the eggNOG category, identifier, conserved PFAM domains, and protein product description shown on the right. For each species and orthologous group, evidence of regulation is shown in green-red shades. Light green shading indicates promoters associated with the ortholog with sites close or equal to the maximum site score detected in the analysis. Red shading denotes promoters in which the best scoring site has a score close or equal to the minimum score observed in the analysis. Blue shading indicates that no orthologous sequences have been identified in a given species for a particular ortholog group.

TIGR03916 (ENOG4105ES7) and TIGR03915 (ENOG4105T12) models. These models are present in about 20% of sequenced bacterial genomes, arranged always in operon configuration, and are thought to constitute a DNA base excision repair system involving a uracil-DNA glycosylase (UDG) domain that is conserved in all *Verrucomicrobia* TIGR03915-matching homologs.

To validate the predictions of the comparative genomics approach and further establish the LexA regulon of the *Verrucomicrobia*, we performed EMSA with purified *V. spinosum* and *O. terrae* proteins on the promoter region of several genes with predicted LexA-binding sites in these organisms and evidence of regulation in at least three different genomes. The results, shown in **Figure 6**, confirm that LexA binds to the promoter region of the *splB* gene in *O. terrae* (OTER_RS07185) and *V. spinosum* (VSP_RS12190). *V. spinosum* LexA also binds the *imuA-imuB-dnaE2* operon promoter, the promoters of genes coding for DNA polymerase IV (VSP_RS08510) and RecQ (VSP_RS32195) homologs, and the *recA* (VSP_RS04780) and *uvrA* (VSP_RS32650) promoters. Together with the comparative genomics analysis, these results confirm the existence of a conserved core LexA-regulon in the *Verrucomicrobia* and demonstrate that, in some *Verrucomicrobia* species, LexA controls a network of similar size and function to those reported in well-studied bacterial phyla, using a novel LexA-binding motif.
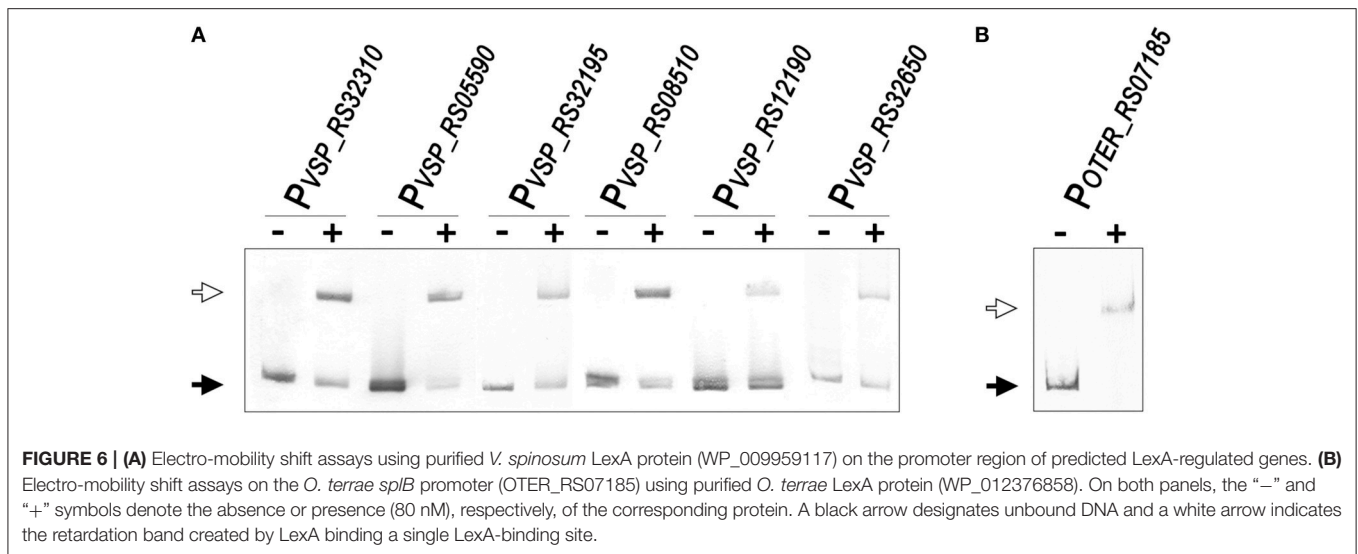
DISCUSSION

This work reports the combined use of *in silico* and *in vitro* techniques to characterize a novel binding motif for the SOS

transcriptional repressor LexA in the *Verrucomicrobia*, and its use to define the LexA regulon in this bacterial phylum of emerging interest. The results provide further context to illustrate the complex evolutionary history of the SOS response, and put to the fore the plasticity and versatility of this transcriptional system.

Variability and Core Elements of the SOS Regulatory Network

Phylogenetic and protein signature analyses have firmly established the *Verrucomicrobia* as one of the major phyla in the PVC super-phylum, an ancient bacterial group estimated to have diverged from other bacterial clades almost two billion years ago (Gupta et al., 2012; Kamneva et al., 2012; Lagkouvardos et al., 2014). The analysis of the LexA regulon performed in this work hence provides for the first time insights into the organization of a complex transcriptional system in this large bacterial clade. Our results show evidence of a functional LexA protein targeting the same LexA-binding motif in all three major *Verrucomicrobia* classes (*Opitutae, Spartobacteria*, and *Verrucomicrobiae*), suggesting that a functional LexA regulatory network was present in the ancestor of the *Verrucomicrobia* (**Figure 1**). However, the *Verrucomicrobia* also display substantial heterogeneity in the size of their predicted LexA regulons (**Figure 5**). Some families, such as the *Methylacidiphilaceae* do not present LexA homologs, while many members of the Verrucomicrobiales display small (1–3 operon) regulons, a setup that appears to be the rule in the Opitutales. Small SOS regulatory networks have been experimentally reported for several species, but mostly in association with drastic changes in the LexA-binding motif

FIGURE 6 | (A) Electro-mobility shift assays using purified *V. spinosum* LexA protein (WP_009959117) on the promoter region of predicted LexA-regulated genes. **(B)** Electro-mobility shift assays on the *O. terrae splB* promoter (OTER_RS07185) using purified *O. terrae* LexA protein (WP_012376858). On both panels, the "−" and "+" symbols denote the absence or presence (80 nM), respectively, of the corresponding protein. A black arrow designates unbound DNA and a white arrow indicates the retardation band created by LexA binding a single LexA-binding site.

(Jara et al., 2003; Campoy et al., 2005; Mazon et al., 2006). In these instances, the LexA regulon is typically constrained to the regulation of translesion synthesis polymerases (Erill et al., 2007). Conversely, moderately large (10–40 genes) LexA regulons incorporating several DNA repair pathways have been documented in the *Gammaproteobacteria*, the *Betaproteobacteria*, the *Actinobacteria,* and the *Firmicutes* (Fernandez De Henestrosa et al., 2000; Davis et al., 2002; Au et al., 2005; Ulrich et al., 2013; Sanchez-Alberola et al., 2015). These findings have substantiated the notion that translesion synthesis is the primordial function of the SOS response, and the identification of a translesion synthesis operon (*imuA-imuB-dnaE2*) in the core LexA regulon of the *Verrucomicrobia* confirms the ancestral role of this mechanism in the SOS response. Nonetheless, the presence of a putative photolyase (*splB*) in the core *Verrucomicrobia* LexA regulon, with documented LexA-regulated orthologs in several bacterial clades (Davis et al., 2002; Cirz et al., 2006; Sanchez-Alberola et al., 2012, 2015; Ulrich et al., 2013), suggests that photoreactivation might have played an essential DNA repair role in the primordial SOS response.

Beyond the presence of a putative photolyase, the SOS response of the *Verrucomicrobia* presents several interesting differences with the canonical SOS response of *E. coli* and *B. subtilis*. The *Opitutae*, for instance, show a consistent absence of LexA regulation for the *recA* gene. The lack of *recA* regulation by LexA has been reported in several bacterial groups, such as the *Acidobacteria* and the *Deltaproteobacteria* (Jara et al., 2003; Campoy et al., 2005; Mazon et al., 2006). Loss of *recA* regulation is often associated with small LexA regulons and *lexA* gene duplication, which are both features of the *Opitutae* LexA network inferred in this work. Another distinct feature of the *Verrucomicrobia* LexA regulon is the regulation of multiple RecQ homologs (ENOG4107QS5 and ENOG410QKP1; **Figure 5**). One of these RecQ homologs (ENOG4107QS5) shares functional domains with *B. subtilis* RecS and RecQ proteins and therefore likely fulfills similar repair functions. The other RecQ homolog

(ENOG410QKP1) lacks DNA-binding HRDC (Helicase and RNase D C-terminal) and RecQ-C-terminal (RQC) domains and presents weaker evidence of homology with *B. subtilis* RecS and RecQ proteins (Fernández et al., 1998). RecQ helicases are involved in the initiation and reversal of recombination and are known to act in concert with the product of SOS genes (*recA*, *ssb*), and to facilitate the onset of the SOS response (Heyer, 2004; Nakayama, 2005). Although SOS regulation of RecQ homologs has not been documented to date, LexA regulation of other DNA helicases (UvrD, PcrA, DinG) is a well-established feature of the SOS response in several organisms (Fernandez De Henestrosa et al., 2000; Au et al., 2005; Abella et al., 2007). These helicases do not appear to be regulated in the *Verrucomicrobia*, suggesting that the putative LexA regulation of RecQ homologs might be fulfilling a complementary role in this phylum. Our analysis also provides evidence of LexA regulation for an operon encoding radical SAM and uracil-DNA glycosylase domain-containing proteins (ENOG4105ES7 and ENOG4105T12; **Figure 5**), presumed to function as a DNA base excision repair system. SOS-regulated error-prone polymerases have been shown to have poor sugar discrimination, leading to the frequent misincorporation of ribonucleotides (Schroeder et al., 2015). Misincorporated ribonucleotides are usually removed by RNase HII-mediated ribonucleotide excision repair, but SOS-regulated nucleotide excision repair has also been shown to address ribonucleotide incorporation (Vaisman et al., 2013). The presence of putative LexA-regulated translesion synthesis polymerases in the *Verrucomicrobia* (**Figures 5, 6**) hence suggests that the regulation of this base excision repair operon by LexA may play a role in addressing uracil misincorporation resulting from SOS induction in this phylum.
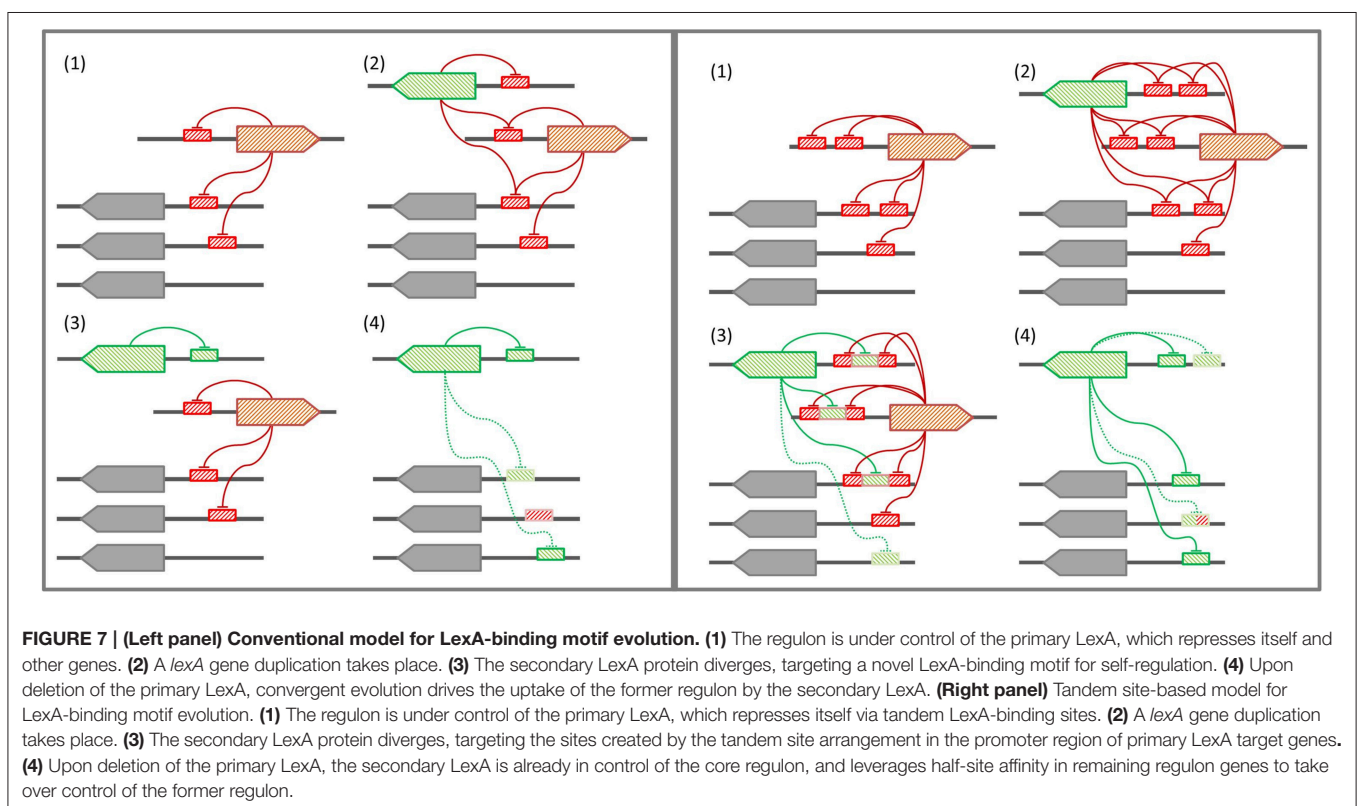
A Tandem Model for the Evolution of the LexA-Binding Motif

In those bacterial phyla where the SOS response has been experimentally documented, the LexA-binding motif shows

evidence of high conservation, punctuated by periods of rapid divergence and further stabilization (Erill et al., 2007). In the *Firmicutes* and the *Actinobacteria*, LexA targets a conserved GAAC-N4-GTTC LexA-binding motif that is monophyletic for both clades (Cornish et al., 2012), and variations of this motif are also seen in other bacterial groups, such as the *Cyanobacteria* or the Chloroflexi (Fernandez de Henestrosa et al., 2002; Mazon et al., 2004b). In the *Proteobacteria*, however, LexA shows an extraordinary diversity of binding motifs (Erill et al., 2007). In *Proteobacteria* classes with abundant sequence information (*Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria*), the LexA-binding motif has been found to be extremely well-conserved, but exceptions to the canonical LexA-binding motif of *Gammaproteobacteria* and *Betaproteobacteria* have been reported in several subgroups (Campoy et al., 2002; Abella et al., 2007; Sanchez-Alberola et al., 2015). These exceptions are associated with duplications of the *lexA* gene, suggesting a model for LexA-binding motif evolution (**Figure 7**) in which *lexA* duplication leads to progressive divergence in the LexA-binding motif of the duplicated *lexA*, until the primary *lexA* gene is deleted and the divergent LexA takes control of the regulon (Abella et al., 2007; Yang et al., 2008; Sanchez-Alberola et al., 2015). While this model provides a causal mechanism for LexA-motif divergence, it does not address how a divergent LexA can swiftly take control over a regulatory network defined, up to the deletion event, by LexA-binding sites matching the primary LexA-binding motif. Furthermore, the model does not provide a mechanistic explanation for the recurrence of very

similar LexA-binding motifs in distantly related bacterial clades, such as the *Firmicutes* and the Gallionellales, recognized through seemingly unrelated LexA α3 helices (Sanchez-Alberola et al., 2015).

Many bacterial transcription factors bind cooperatively to tandem sites (Barnard et al., 2004). The existence of tandem sites for LexA was first reported in the promoter region of the *E. coli lexA* gene (Brent, 1982) and then shown to be a common feature of *lexA* genes in the *Gammaproteobacteria* (Garriga et al., 1992), the *Betaproteobacteria*, and *Alphaproteobacteria* (Sanchez-Alberola et al., 2012), and the *Firmicutes* and *Actinobacteria* (Cornish et al., 2012). These arrangements feature highly conserved and spatially close tandem sites (1–10 bp apart). Tandem LexA-binding sites have also been experimentally reported for other SOS genes, such as the *ydjM* gene of *E. coli* (Fernandez De Henestrosa et al., 2000) or the *umuDC*-like operon (*yqjW-yqzH*) of *B. subtilis* (Au et al., 2005). Furthermore, the use of cooperative LexA-binding to enhance repression has been experimentally demonstrated for several colicin genes, which display a tandem arrangement with a strong and a weak LexA-binding site overlapping at their terminal positions (Gillor et al., 2008). In the *Verrucomicrobia*, there is evidence of a recent *lexA* duplication in the *Opituctaceae* and tandem LexA-binding sites separated by short distances appear to be a conserved feature of the *lexA* promoter (**Figure 3**). The ability of the *Verrucomicrobia* LexA to cooperatively bind degenerate sites and the fact that at least in one of these species the tandem arrangement generates a functional *B. subtilis* LexA-binding

FIGURE 7 | (Left panel) Conventional model for LexA-binding motif evolution. (1) The regulon is under control of the primary LexA, which represses itself and other genes. **(2)** A *lexA* gene duplication takes place. **(3)** The secondary LexA protein diverges, targeting a novel LexA-binding motif for self-regulation. **(4)** Upon deletion of the primary LexA, convergent evolution drives the uptake of the former regulon by the secondary LexA. **(Right panel)** Tandem site-based model for LexA-binding motif evolution. **(1)** The regulon is under control of the primary LexA, which represses itself via tandem LexA-binding sites. **(2)** A *lexA* gene duplication takes place. **(3)** The secondary LexA protein diverges, targeting the sites created by the tandem site arrangement in the promoter region of primary LexA target genes**. (4)** Upon deletion of the primary LexA, the secondary LexA is already in control of the core regulon, and leverages half-site affinity in remaining regulon genes to take over control of the former regulon.

motif (**Figure 4**) suggest that tandem site arrangements can yield a simple mechanistic process for the evolution of LexA-binding motifs.

In the tandem site model (**Figure 7**), LexA binds consecutive sites in its own promoter and in the promoter of key SOS genes that need to be tightly regulated (Gillor et al., 2008). Upon *lexA* duplication, the site generated by the tandem arrangement provides the secondary LexA with a conserved target for motif divergence. This allows the secondary LexA to maintain cross-regulation with the primary LexA and a subset of its regulon, while incorporating novel elements to its network. After deletion of the primary *lexA* gene, the secondary LexA is hence already in control of a core LexA regulon, and can rapidly evolve sites on other target genes by exploiting its partial overlap, and presumable weak binding affinity, with primary *lexA* sites. The tandem site model therefore provides a conservative mechanism for the evolution of LexA-binding motifs that is capable of addressing outstanding questions regarding the complex evolutionary history of the SOS response. On the one hand, the conservative nature of the model provides a natural explanation for the persistence of conserved SOS response networks under divergent LexA-binding motifs, without the need for a strong selective process driving convergent evolution of similar networks (Sanchez-Alberola et al., 2015). On the other hand, the implicit reuse of LexA monomer-binding sites in the tandem model helps explain the observation of many LexA-binding motifs involving the rearrangement of similar monomer binding sites on different motif structures (Mazon et al., 2004a; Sanchez-Alberola et al., 2015).

Several lines of evidence provide indirect support for a tandem site-based model of LexA-binding motif evolution. As mentioned above, the prevalence of such arrangements in the promoter of *lexA* and other SOS genes has been documented in several bacterial clades. Furthermore, *lexA* duplications targeting identical and divergent motifs have also been experimentally reported, and cross-regulation between duplicated *lexA* genes has been demonstrated in these systems (Jara et al., 2003; Abella et al., 2007; Yang et al., 2008). Lastly, previous work has shown that LexA can bind to degenerate sites that partially match other LexA-binding motifs, indicating that transitional stages of LexA divergence in which the secondary LexA could partially bind the original and tandem-generated motifs are possible (Mazon et al., 2004a). Due to its broad distribution in several phyla, the *Firmicutes* and *Actinobacteria* LexA-binding motif has long been assumed to represent the ancestral motif of LexA. The mirror image relationship between *Firmicutes* and *Verrucomicrobia* LexA-binding motifs, and the generation of functional *B. subtilis* LexA-binding sites from tandem *Verrucomicrobia* LexA-binding sites, hence suggest that the *Verrucomicrobia* LexA-binding motif might have originated after the duplication of a *lexA* gene targeting a tandem arrangement of *Firmicutes*-like LexA-binding sites in a common ancestor of these lineages. The analysis of the α3 helix of the *Verrucomicrobia* LexA DNA-binding domain (**Figure 2**) supports this hypothesis, revealing overall conservation of amino acid properties and a substitution pattern consistent with changes in the specific readout of monomer sites, but not in overall motif recognition, as expected in the tandem site evolution model.

CONCLUSIONS

By combining *in silico* and *in vitro* methods, in this work we have characterized a novel LexA-binding motif in the *Verrucomicrobia*. Using this motif, which presents structural similarities with LexA-binding motifs previously described in other phyla, we performed a comparative genomics analysis of the LexA regulon in this understudied phylum. Our computational analysis, validated through *in vitro* assays, revealed significant variability in the size and composition of the LexA regulatory network of this phylum, and identified novel core and ancillary components of the SOS response. The characterization of the *Verrucomicrobia* LexA-binding motif and regulon also allowed us to postulate for the first time a model for LexA-binding motif evolution that satisfactorily addresses open questions in the evolution of this system via gene duplication events. Future biochemical and genetic experiments, such as determining the conformation of LexA in solution and analyzing expression patterns in mutants for core SOS genes, should provide a more comprehensive characterization of the *Verrucomicrobia* SOS response and its evolution.

AUTHOR CONTRIBUTIONS

IE and SK performed the *in silico* analyses. SK developed scripts for comparative genomics. SC performed the *in vitro* analyses. IE, SC, and JB discussed the findings and interpreted the results. IE and JB conceived the experiment and coordinated the research. IE drafted the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmolb.2016.00033

REFERENCES

Abella, M., Campoy, S., Erill, I., Rojo, F., and Barbe, J. (2007). Cohabitation of two different lexA regulons in *Pseudomonas putida*. *J. Bacteriol.* 189, 8855–8862. doi: 10.1128/JB.01213-07

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Au, N., Kuester-Schoeck, E., Mandava, V., Bothwell, L. E., Canny, S. P., Chachu, K., et al. (2005). Genetic composition of the *Bacillus subtilis* SOS system. *J. Bacteriol.* 187, 7655–7666. doi: 10.1128/JB.187.22.7655-7666.2005

Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–W373. doi: 10.1093/nar/gkl198

Barnard, A., Wolfe, A., and Busby, S. (2004). Regulation at complex bacterial promoters: how *bacteria* use different promoter organizations to produce different regulatory outcomes. *Curr. Opin. Microbiol.* 7, 102–108. doi: 10.1016/j.mib.2004.02.011

Beaber, J. W., Hochhut, B., and Waldor, M. K. (2004). SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* 427, 72–74. doi: 10.1038/nature02241

Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, J. G., Walters, W. A., et al. (2011). The under-recognized dominance of *Verrucomicrobia* in soil bacterial communities. *Soil Biol. Biochem.* 43, 1450–1455. doi: 10.1016/j.soilbio.2011.03.012

Bhargava, N., and Erill, I. (2010). xFITOM: a generic GUI tool to search for transcription factor binding sites. *Bioinformation* 5, 49–50. doi: 10.6026/97320630005049

Brent, R. (1982). Regulation and autoregulation by LexA protein. *Biochimie* 64, 565–569. doi: 10.1016/S0300-9084(82)80088-6

Cambray, G., Sanchez-Alberola, N., Campoy, S., Guerin, E., Da Re, S., Gonzalez-Zorn, B., et al. (2011). Prevalence of SOS-mediated control of integron integrase expression as an adaptive trait of chromosomal and mobile integrons. *Mob. DNA* 2:6. doi: 10.1186/1759-8753-2-6

Campoy, S., Mazon, G., Fernandez de Henestrosa, A. R., Llagostera, M., Monteiro, P. B., and Barbe, 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. doi: 10.1099/00221287-148-11-3583

Campoy, S., Salvador, N., Cortes, P., Erill, I., and Barbe, J. (2005). Expression of canonical SOS genes is not under LexA repression in *Bdellovibrio bacteriovorus*. *J. Bacteriol.* 187, 5367–5375. doi: 10.1128/JB.187.15.5367-5375.2005

Cirz, R. T., O'Neill, B. M., Hammond, J. A., Head, S. R., and Romesberg, F. E. (2006). Defining the *Pseudomonas aeruginosa* SOS response and its role in the global response to the antibiotic ciprofloxacin. *J. Bacteriol.* 188, 7101–7110. doi: 10.1128/JB.00807-06

Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* 6, 786–787. doi: 10.1038/nmeth1109-786

Cornish, J. P., Matthews, F., Thomas, J. R., and Erill, I. (2012). Inference of self-regulated transcriptional networks by comparative genomics. *Evol. Bioinform. Online* 8, 449–461. doi: 10.4137/EBO.S9205

Cornish, J. P., Sanchez-Alberola, N., O'Neill, P. K., O'Keefe, R., Gheba, J., and Erill, I. (2014). Characterization of the SOS meta-regulon in the human gut microbiome. *Bioinformatics* 30, 1193–1197. doi: 10.1093/bioinformatics/btt753

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004

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. doi: 10.1128/JB.184.12.3287-3295.2002

Dubourg, G., Lagier, J.-C., Armougom, F., Robert, C., Audoly, G., Papazian, L., et al. (2013). High-level colonisation of the human gut by *Verrucomicrobia* following broad-spectrum antibiotic treatment. *Int. J. Antimicrob. Agents* 41, 149–155. doi: 10.1016/j.ijantimicag.2012.10.012

Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

Erill, I., Campoy, S., and Barbe, J. (2007). Aeons of distress: an evolutionary perspective on the bacterial SOS response. *FEMS Microbiol. Rev.* 31, 637–656. doi: 10.1111/j.1574-6976.2007.00082.x

Erill, I., Campoy, S., Mazon, G., and Barbe, J. (2006). Dispersal and regulation of an adaptive mutagenesis cassette in the bacteria domain. *Nucleic Acids Res.* 34, 66–77. doi: 10.1093/nar/gkj412

Erill, I., Escribano, M., Campoy, S., and Barbe, J. (2003). *In silico* analysis reveals substantial variability in the gene contents of the *gamma proteobacteria* LexA-regulon. *Bioinformatics* 19, 2225–2236. doi: 10.1093/bioinformatics/btg303

Erill, I., Jara, M., Salvador, N., Escribano, M., Campoy, S., and Barbe, J. (2004). Differences in LexA regulon structure among *Proteobacteria* through *in vivo* assisted comparative genomics. *Nucleic Acids Res.* 32, 6617–6626. doi: 10.1093/nar/gkh996

Federhen, S. (2012). The NCBI Taxonomy database. *Nucleic Acids Res.* 40, D136–D143. doi: 10.1093/nar/gkr1178

Fernandez de Henestrosa, A. R., Cune, J., Erill, I., Magnuson, J. K., and Barbe, J. (2002). A green nonsulfur bacterium, *Dehalococcoides* ethenogenes, with the LexA binding sequence found in gram-positive organisms. *J. Bacteriol.* 184, 6073–6080. doi: 10.1128/JB.184.21.6073-6080.2002

Fernandez De Henestrosa, A. R., Ogi, T., Aoyagi, S., Chafin, D., Hayes, J. J., Ohmori, H., et al. (2000). Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. *Mol. Microbiol.* 35, 1560–1572. doi: 10.1046/j.1365-2958.2000.01826.x

Fernandez de Henestrosa, A. R., Rivera, E., Tapias, A., and Barbe, J. (1998). Identification of the *Rhodobacter sphaeroides* SOS box. *Mol. Microbiol.* 28, 991–1003. doi: 10.1046/j.1365-2958.1998.00860.x

Fernández, S., Sorokin, A., and Alonso, J. C. (1998). Genetic recombination in *Bacillus subtilis* 168: effects of recU and recS mutations on DNA repair and homologous recombination. *J. Bacteriol.* 180, 3405–3409.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Garriga, X., Calero, S., and Barbe, J. (1992). Nucleotide sequence analysis and comparison of the *lexA* genes from *Salmonella* typhimurium, *Erwinia* carotovora, *Pseudomonas aeruginosa* and *Pseudomonas putida*. *Mol. Gen. Genet.* 236, 125–134.

Garrity, G. M., and Holt, J. G. (2001). "The road map to the manual," in *Bergey's Manual® of Systematic Bacteriology*, eds D. R. Boone, R. W. Castenholz, and G. M. Garrity (New York, NY: Springer), 119–166. Available online at: //link.springer.com/chapter/10.1007/978-0-387-21609-6_15 (Accessed May 12, 2016).

Gillor, O., Vriezen, J. A. C., and Riley, M. A. (2008). The role of SOS boxes in enteric bacteriocin regulation. *Microbiol. Read. Engl.* 154, 1783–1792. doi: 10.1099/mic.0.2007/016139-0

Green, M. R., and Sambrook, J. (2012). *Molecular Cloning: A Laboratory Manual (Fourth Edition): Three-volume set*, 4th Edn. New York, NY: Cold Spring Harbor Laboratory Press.

Groban, E. S., Johnson, M. B., Banky, P., Burnett, P. G., Calderon, G. L., Dwyer, E. C., et al. (2005). Binding of the *Bacillus subtilis* LexA protein to the SOS operator. *Nucleic Acids Res.* 33, 6287–6295. doi: 10.1093/nar/gki939

Guerin, E., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Da Re, S., et al. (2009). The SOS response controls integron recombination. *Science* 324, 1034. doi: 10.1126/science.1172914

Gupta, R. S., Bhandari, V., and Naushad, H. S. (2012). Molecular signatures for the PVC Clade (*Planctomycetes*, *Verrucomicrobia*, *Chlamydiae*, and *Lentisphaerae*) of *Bacteria* provide insights into their evolutionary relationships. *Front. Microbiol.* 3:327. doi: 10.3389/fmicb.2012.00327

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., and Noble, W. S. (2007). Quantifying similarity between motifs. *Genome Biol.* 8:R24. doi: 10.1186/gb-2007-8-2-r24

Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., and Beck, E. (2013). TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* 41, D387–D395. doi: 10.1093/nar/gks1234

Heyer, W. D. (2004). Damage signaling: RecQ sends an SOS to you. *Curr. Biol.* 14, R895–R897. doi: 10.1016/j.cub.2004.09.073

Jara, M., Nunez, C., Campoy, S., Fernandez de Henestrosa, A. R., Lovley, D. R., and Barbe, J. (2003). Geobacter sulfurreducens has two autoregulated lexA genes whose products do not bind the recA promoter: differing responses of lexA and

recA to DNA damage. *J. Bacteriol.* 185, 2493–2502. doi: 10.1128/JB.185.8.2493-2502.2003

Kamneva, O. K., Knight, S. J., Liberles, D. A., and Ward, N. L. (2012). Analysis of genome content evolution in PVC bacterial super-phylum: assessment of candidate genes associated with cellular organization and lifestyle. *Genome Biol. Evol.* 4, 1375–1390. doi: 10.1093/gbe/evs113

Kiliç, S., White, E. R., Sagitova, D. M., Cornish, J. P., and Erill, I. (2014). CollecTF: a database of experimentally validated transcription factor-binding sites in *Bacteria*. *Nucleic Acids Res.* 42, D156–D160. doi: 10.1093/nar/gkt1123

Lagkouvardos, I., Jehl, M. A., Rattei, T., and Horn, M. (2014). Signature protein of the PVC superphylum. *Appl. Environ. Microbiol.* 80, 440–445. doi: 10.1128/AEM.02655-13

Liou, A. P., Paziuk, M., Luevano, J.-M., Machineni, S., Turnbaugh, P. J., and Kaplan, L. M. (2013). Conserved shifts in the gut microbiota due to gastric bypass reduce host weight and adiposity. *Sci. Transl. Med.* 5, 178ra41. doi: 10.1126/scitranslmed.3005687

Little, J. W. (1991). Mechanism of specific LexA cleavage: autodigestion and the role of RecA coprotease. *Biochimie* 73, 411–421. doi: 10.1016/0300-9084(91)90108-D

Little, J. W., and Mount, D. W. (1982). The SOS regulatory system of *Escherichia coli*. *Cell* 29, 11–22. doi: 10.1016/0092-8674(82)90085-X

Maiques, E., Ubeda, C., Campoy, S., Salvador, N., Lasa, I., Novick, R. P., et al. (2006). beta-lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus* aureus. *J. Bacteriol.* 188, 2726–2729. doi: 10.1128/JB.188.7.2726-2729.2006

Markowitz, V. M., Chen, I.-M. A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., et al. (2014). IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* 42, D568–D573. doi: 10.1093/nar/gkt919

Mazon, G., Campoy, S., Erill, I., and Barbe, J. (2006). Identification of the *Acidobacterium capsulatum* LexA box reveals a lateral acquisition of the *Alphaproteobacteria lexA* gene. *Microbiology* 152, 1109–1118. doi: 10.1099/mic.0.28376-0

Mazon, G., Erill, I., Campoy, S., Cortes, P., Forano, E., and Barbe, J. (2004a). Reconstruction of the evolutionary history of the LexA-binding sequence. *Microbiology* 150, 3783–3795. doi: 10.1099/mic.0.27315-0

Mazon, G., Lucena, J. M., Campoy, S., Fernandez de Henestrosa, A. R., Candau, P., and Barbe, J. (2004b). LexA-binding sequences in Gram-positive and cyanobacteria are closely related. *Mol. Genet. Genomics* 271, 40–49. doi: 10.1007/s00438-003-0952-x

Nakayama, H. (2005). *Escherichia coli* RecQ helicase: a player in thymineless death. *Mutat. Res. Mol. Mech. Mutagen.* 577, 228–236. doi: 10.1016/j.mrfmmm.2005.02.015

Oertel-Buchheit, P., Lamerichs, R. M., Schnarr, M., and Granger-Schnarr, M. (1990). Genetic analysis of the LexA repressor: isolation and characterization of LexA(Def) mutant proteins. *Mol. Gen. Genet.* 223, 40–48. doi: 10.1007/BF00315795

Ottleben, G., Messori, L., Rüterjans, H., Kaptein, R., Granger-Schnarr, M., and Schnarr, M. (1991). 1H-NMR investigation of the interaction of the amino terminal domain of the LexA repressor with a synthetic half-operator. *J. Biomol. Struct. Dyn.* 9, 447–461. doi: 10.1080/07391102.1991.10507928

Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., et al. (2014). eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.* 42, D231–D239. doi: 10.1093/nar/gkt1253

Pruitt, K. D., Tatusova, T., Brown, G. R., and Maglott, D. R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* 40, D130–D135. doi: 10.1093/nar/gkr1079

Sait, M., Kamneva, O. K., Fay, D. S., Kirienko, N. V., Polek, J., Shirasu-Hiza, M. M., et al. (2011). Genomic and experimental evidence suggests that *Verrucomicrobium spinosum* Interacts with *Eukaryotes*. *Front. Microbiol.* 2:211. doi: 10.3389/fmicb.2011.00211

Sanchez-Alberola, N., Campoy, S., Barbe, J., and Erill, I. (2012). Analysis of the SOS response of *Vibrio* and other *bacteria* with multiple chromosomes. *BMC Genomics* 13:58. doi: 10.1186/1471-2164-13-58

Sanchez-Alberola, N., Campoy, S., Emerson, D., Barbe, J., and Erill, I. (2015). A SOS regulon under control of a non-canonical LexA-binding motif in the *Betaproteobacteria*. *J. Bacteriol.* 197, 2622–2630. doi: 10.1128/JB.00035-15

Sassanfar, M., and Roberts, J. W. (1990). Nature of the SOS-inducing signal in *Escherichia coli*. The involvement of DNA replication. *J. Mol. Biol.* 212, 79–96. doi: 10.1016/0022-2836(90)90306-7

Schlieper, D., Oliva, M. A., Andreu, J. M., and Löwe, J. (2005). Structure of bacterial tubulin BtubA/B: evidence for horizontal gene transfer. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9170–9175. doi: 10.1073/pnas.0502859102

Schneider, T. D. (1997). Information content of individual genetic sequences. *J. Theor. Biol.* 189, 427–441. doi: 10.1006/jtbi.1997.0540

Schroeder, J. W., Randall, J. R., Matthews, L. A., and Simmons, L. A. (2015). Ribonucleotides in bacterial DNA. *Crit. Rev. Biochem. Mol. Biol.* 50, 181–193. doi: 10.3109/10409238.2014.981647

Söding, J., Biegert, A., and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–W248. doi: 10.1093/nar/gki408

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33

Thliveris, A. T., Little, J. W., and Mount, D. W. (1991). Repression of the *E coli* recA gene requires at least two LexA protein monomers. *Biochimie* 73, 449–456. doi: 10.1016/0300-9084(91)90112-E

Thliveris, A. T., and Mount, D. W. (1992). Genetic identification of the DNA binding domain of *Escherichia coli* LexA protein. *Proc. Natl. Acad. Sci. U.S.A.* 89, 4500–4504. doi: 10.1073/pnas.89.10.4500

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Ubeda, C., Maiques, E., Knecht, E., Lasa, I., Novick, R. P., and Penades, J. R. (2005). Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. *Mol. Microbiol.* 56, 836–844. doi: 10.1111/j.1365-2958.2005.04584.x

Ubeda, C., Maiques, E., Tormo, M. A., Campoy, S., Lasa, I., Barbe, J., et al. (2007). SaPI operon I is required for SaPI packaging and is controlled by LexA. *Mol. Microbiol.* 65, 41–50. doi: 10.1111/j.1365-2958.2007.05758.x

Ulrich, R. L., Deshazer, D., Kenny, T. A., Ulrich, M. P., Moravusova, A., Opperman, T., et al. (2013). Characterization of the *Burkholderia thailandensis* SOS response by using whole-transcriptome shotgun sequencing. *Appl. Environ. Microbiol.* 79, 5830–5843. doi: 10.1128/AEM.00538-13

Vaisman, A., McDonald, J. P., Huston, D., Kuban, W., Liu, L., Houten, B. V., et al. (2013). Removal of misincorporated ribonucleotides from prokaryotic genomes: an unexpected role for nucleotide excision repair. *PLoS Genet.* 9:e1003878. doi: 10.1371/journal.pgen.1003878

Walker, G. C., Smith, B. T., and Sutton, M. D. (2000). "The SOS response to DNA damage," in *Bacterial Stress Responses*, eds G. Storz and R. Hengge-Aronis (Washington, DC: ASM Press), 131–144.

Yang, M. K., Hsu, C. H., and Sung, V. L. (2008). Analyses of binding sequences of the two LexA proteins of *Xanthomonas* axonopodis pathovar citri. *Mol. Genet. Genomics* 280, 49–58. doi: 10.1007/s00438-008-0344-3

Zhang, A. P. P., Pigli, Y. Z., and Rice, P. A. (2010). Structure of the LexA-DNA complex and implications for SOS box measurement. *Nature* 466, 883–886. doi: 10.1038/nature09200