Analysis of the SOS response of *Vibrio* and other bacteria with multiple chromosomes

Neus Sanchez-Alberola1,2, Susana Campoy1, Jordi Barbé1 and Ivan Erill2*

**Abstract**

**Background:** The SOS response is a well-known regulatory network present in most bacteria and aimed at addressing DNA damage. It has also been linked extensively to stress-induced mutagenesis, virulence and the emergence and dissemination of antibiotic resistance determinants. Recently, the SOS response has been shown to regulate the activity of integrases in the chromosomal superintegrons of the *Vibrionaceae*, which encompasses a wide range of pathogenic species harboring multiple chromosomes. Here we combine *in silico* and *in vitro* techniques to perform a comparative genomics analysis of the SOS regulon in the *Vibrionaceae*, and we extend the methodology to map this transcriptional network in other bacterial species harboring multiple chromosomes.

**Results:** Our analysis provides the first comprehensive description of the SOS response in a family (*Vibrionaceae*) that includes major human pathogens. It also identifies several previously unreported members of the SOS transcriptional network, including two proteins of unknown function. The analysis of the SOS response in other bacterial species with multiple chromosomes uncovers additional regulon members and reveals that there is a conserved core of SOS genes, and that specialized additions to this basic network take place in different phylogenetic groups. Our results also indicate that across all groups the main elements of the SOS response are always found in the large chromosome, whereas specialized additions are found in the smaller chromosomes and plasmids.

**Conclusions:** Our findings confirm that the SOS response of the *Vibrionaceae* is strongly linked with pathogenicity and dissemination of antibiotic resistance, and suggest that the characterization of the newly identified members of this regulon could provide key insights into the pathogenesis of *Vibrio*. The persistent location of key SOS genes in the large chromosome across several bacterial groups confirms that the SOS response plays an essential role in these organisms and sheds light into the mechanisms of evolution of global transcriptional networks involved in adaptability and rapid response to environmental changes, suggesting that small chromosomes may act as evolutionary test beds for the rewiring of transcriptional networks.
repressors, such as the λ phage CI repressor. Cleaved LexA is unable to bind its target sites, thus activating the expression of SOS genes [9]. The E. coli SOS transcriptional network comprises more than 40 genes involved in diverse forms of DNA repair, error-prone DNA replication and cell division arrest, and it includes also both the recA and lexA genes [5,10,11]. Once DNA damage has been addressed, activated RecA levels fall rapidly and newly synthesized LexA returns the system to its repressed state.

The SOS response was first identified in response to UV treatment, but it is also intrinsically linked to the natural response against antibiotics that damage DNA [12]. Furthermore, the need to bypass lesions that cannot be readily repaired using translesion synthesis polymerases (TLS) establishes the SOS response as a de facto mechanism for stress-induced mutagenesis [13-15]. In addition, the SOS response has long been known to promote prophage induction [16] and the activation of virulence factors, like Shiga-like toxins in enterohemorrhagic E. coli [17]. Independent studies have shown that activation of the SOS response also triggers the dissemination of other mobile elements, such as pathogenicity islands [18]. Finally, it has been reported that several antibiotics that do not cause direct damage to bacterial DNA (e.g. β-lactams) are capable of inducing the SOS response through indirect routes [19], and that such induction is capable of triggering dissemination of mobile elements carrying antibiotic resistance determinants [20]. These findings have irrevocably transformed the image of the SOS response from a DNA-repair textbook paradigm into a key component in the fight against virulence and antibiotic resistance [21-23].

The recent evidence for a tight link between dissemination of antibiotic resistance, pathogenicity and the SOS response is perhaps most clearly outlined in the genus Vibrio. This group of heterotrophic and aerobic γ-Proteobacteria encompasses at least 12 species known for being pathogenic in human hosts [24]. Vibrio species are usually linked to intestinal infections (e.g. *Vibrio cholerae* or *Vibrio parahaemolyticus*) acquired by drinking of contaminated water or eating of raw sea animals [25-27], but they are also well documented as opportunistic pathogens associated with high morbidity rates (e.g. *Vibrio alginolyticus* or *Vibrio vulnificus*) [28-30]. The genome of *Vibrio* species is composed of 2 circular chromosomes (known as large/1 and small/2) and in most cases it is known to harbor several virulence determinants, including toxins and colonization factors, that are strongly linked to the well-known severity and harshness of classical *Vibrio* infections, such as cholera [31-33]. In 2004, Beaber *et al.* showed that transfer of the *V. cholerae* SXT integrative-conjugative element (ICE) was controlled by the SOS response through RecA-dependent cleavage of the SXT SetR repressor [34]. The SXT is a 100 kbp element encoding genes that confer resistance to several antibiotics, such as chloramphenicol, sulphonmethoxazole or trimethoprim, and is now found in almost all clinical isolates of *V. cholerae* from Asia and Africa [35]. The fact that SOS-mediated activation of SXT conjugation could be induced by antibiotics, such as ciprofloxacin, put forward for the first time a mechanism by which the use of antibiotic agents might promote the dissemination of antibiotic resistance determinants [34]. Soon thereafter, Quinones *et al.* showed that induction of the CTX prophage, which encodes the cholera toxin, is regulated by the SOS repressor LexA in conjunction with the phage repressor RstR, and the same group recently proved that transcription of the rstR gene is also regulated by LexA [36,37]. Most recently, the SOS response has been shown to regulate the activity of the integrase gene of the *V. cholerae* superintegron, leading to heightened rates of recombination excision of integron cassettes upon induction of the SOS response by different antibiotics [38]. The *V. cholerae* superintegron encompasses more than 150 gene cassettes and is known to be a source of genetic diversity and antibiotic resistance genes [39,40]. SOS regulation of the integrase gene has later been confirmed in chromosomal superintegrons of the *Vibrio* genus and in most mobile integron classes, establishing a mechanism for the long-term safe storage of otherwise harmful adaptive traits, such as antibiotic resistance determinants [41].

Since its initial description in *E. coli*, the SOS response has been described in many other bacteria and, with some notable exceptions (e.g. the γ-Proteobacteria), it is considered a universal stress response in the Bacteria domain [23]. In contrast to other transcriptional networks, like the *arg* regulon [42], the SOS response has undergone extensive evolutionary changes in the binding motif recognized by the repressor protein LexA. As a consequence, the original description of a CTGT-N₈-ACAG binding motif in *E. coli* has been extended to most β- and γ-Proteobacteria species [43,44], but has been replaced by alternative descriptions in many different classes and phyla. The universality of the *E. coli* LexA-binding motif was first challenged by the description of a GAAC-N₇-GTTC LexA-binding motif in the Firmicutes *Bacillus subtilis* [45]. This new motif was later confirmed to be the LexA-binding motif in the Actinobacterium *Mycobacterium tuberculosis* and, by extension, in all Gram-positive bacteria [46]. Following this initial discovery, research has unearthed a plethora of LexA-binding motifs in different bacterial groups, ranging from a GTTC-N₈-GTTC direct repeat in the α-Proteobacteria to a TTAC-N₇-GTAA palindrome in *Bdellovibrio bacteriovorus* [47,48]. The identification of
the LexA-binding motif in several bacterial groups has led to the outlining of a loosely conserved core of genes that seem to constitute the core or ancestral LexA regulon. These include the two main SOS genes (lexA and recA), members of the nucleotide excision repair (NER) system (most notably the excinuclease subunit A, uvrA) and several TLS polymerases [23]. Recently, the canonical view of the SOS response as a DNA repair system has shifted considerably towards a mutagenesis-centered perspective following the identification of a mutagenesis cassette (imuA-imuB-dnaE2) that is widely distributed in the Bacteria domain. Whenever present, this cassette is persistently regulated by LexA, being the only transcriptional unit regulated by the SOS repressor in several organisms [15,48-50].

The rapid increase in the number of completely sequenced bacterial genome sequences in the early 2000s led to the development of comparative genomics approaches for the analysis of transcriptional networks [42,51-53]. The main premise in these approaches is that binding sites for a transcription factor, if functional, will tend to be preserved by evolution. Hence, the analysis of multiple genomes will identify functional binding sites as conserved regions across several genomes, in spite of significant changes to their immediate neighborhood. Functional conservation of binding sites can be exploited in two different ways. In motif discovery applications, the upstream regions of orthologous genes can be scanned to identify consistent instances of a candidate binding motif [54-56]. In contrast, transcriptional regulatory network analyses use a known binding motif to search multiple genomes and elucidate the composition of the regulatory network at different levels [42,57,58]. In this approach, the identification of a candidate binding site upstream of an ortholog in different species bolsters the a priori low confidence of the in silico prediction made on each individual genome [59,60].

A basic tenet of comparative genomic approaches for regulon analysis is the assumption that the transcription factor-binding motif is conserved across the species being studied. In spite of the apparent diversity in LexA-binding motifs across the Bacteria domain, research on different phyla has also shown that changes in LexA-binding motif tend to be monophyletic for the clades in which they take place [61-63]. This has enabled research into the composition and evolution of the SOS response in several bacterial groups using comparative genomics approaches. These studies have revealed that, beyond the conserved core outlined above, the genetic composition of the SOS response can vary substantially between and within Proteobacteria classes [23,43,64]. Because the original collection of binding sites may not belong to any of the species under analysis, a good practice in comparative genomics involves the experimental validation of in silico predictions in at least one of the organisms under study, in order to validate the computational method and the parameters used for the analysis [64].

In this work, we take advantage of previously developed methodology to map for the first time the SOS regulatory network in the Vibrionaceae family. We apply computational analyses to the 14 complete genome sequences available for this family, which encompass nine different Vibrionaceae species, and we validate our findings in vitro using V. parahaemolyticus as an experimental model. Our analysis thus provides a first glimpse at the genetic organization of the SOS response in a complex and heterogeneous family of bacteria comprising more than a hundred species. In order to gain insight into the adaptation of the SOS response to genomes containing multiple chromosomes, which are a defining feature of the Vibrionaceae, we extend the comparative genomics analysis to two additional phylogenetic groups (the β- and α-Proteobacteria) that contain many species with multiple chromosomes. Our results provide comprehensive support to the notion that the SOS response is strongly associated with dissemination of antibiotic resistance and pathogenicity in the Vibrio genus, and identify new genes that might play a role in this association. We also show that the core constituents of the SOS response in all the groups analyzed are located systematically in the large chromosome, indicating that the SOS response is a fundamental component of these genomes. The observed distribution of non-core genes among the smaller chromosomes and plasmids puts forward a mechanism for the dynamic evolution of a critical response network involved in the regulation of many essential genes, and supports the hypothesis of plasmid origin for the smaller chromosomes of organisms with multiple chromosome genomes.

**Results and Discussion**

The SOS regulon of the Vibrionaceae encompasses more than 20 genes

To analyze the SOS regulon of the Vibrionaceae, we downloaded all the available complete genome sequences from the NCBI RefSeq database [65] and we performed independent searches of both chromosomes and sequenced plasmids using a collection of known binding sites from *E. coli* [43]. As in previous work, to define the family-wide SOS regulon we included only putative sites that were observed consistently upstream of orthologous genes in at least three separate species. The results of the comparative genomics analysis shown in Figure 1 suggest that, in accordance with other LexA regulons, the SOS response of the Vibrionaceae
comprises more than 20 genes [5,66]. Given the relative phylogenetic proximity between the Vibrionaceae and E. coli, it is not surprising to find that many of the genes described as SOS regulated in E. coli are also members of the Vibrionaceae SOS regulon. These include the two main regulators of the SOS network (lexA and recA), as well as canonical SOS genes involved in DNA repair (recN, yebG, yigN, uvrA, uvrD and the ruvAB operon) and translesion synthesis (dinB and umuC). Nonetheless, differences in the mode of regulation of certain genes can be observed in the Vibrionaceae. The uvrA gene forms a divergent gene pair with ssb in E. coli and other γ-Proteobacteria, sharing a single LexA-binding site [43]. Expression of both genes has been shown to be SOS regulated, even though SOS induction of ssb is known to be moderate [67,68]. This arrangement (Figure 2A) is preserved in many Vibrionaceae (e.g. V. parahaemolyticus), but has been disrupted by a bacteriophage insertion in other species (e.g. Vibrio fischeri). The fact that independent LexA-binding sites for ssb and uvrA have been preserved in these species points to a selective advantage in the regulation of both genes by the SOS response. In contrast, the ruvCAB operon yields a different story (Figure 2B). In E. coli and close relatives, this operon is split into the ruvC gene and the ruvAB operon, and only the latter is regulated

---

**Figure 1** Tabulated description of the LexA regulon of the Vibrionaceae. Colors indicate the presence and location of the gene and patterns denote presence (plain) or absence (patterned) of one or more LexA-binding sites in its promoter region. E. coli genes and their corresponding regulation are shown for comparative purposes. Abbreviations are as follows: Eco, E. coli; Vch, V. cholerae; Vha, V. harveyi; Vpp, V. parahaemolyticus; Vsp, Vibrio splendidus; Vvu, V. vulnificus; Asa, A. salmonicida; Vfi, V. fischeri; Ppr, Photobacterium profundum.
by the SOS response [5,43]. In Aliivibrio salmonicida and other species, the operon retains its ancestral ruvCAB configuration but displays independent LexA-binding sites upstream of both ruvC and the ruvAB tandem, which presumably incorporates a secondary promoter. The operon has later been disrupted by the insertion of one (V. vulnificus) or two (V. parahaemolyticus) genes and the regulation of ruvC has apparently been lost in the process. This suggests that the selective pressure for SOS regulation of ruvC is not as strong as the one operating on ruvAB, in agreement with recent results that emphasize the independent and vital role of the ruvAB complex in rescuing stalled replication forks [69].

The comparative genomics analysis also identifies putative binding sites upstream of several genes that are not members of the E. coli LexA regulon but that have been shown experimentally to be SOS regulated in at least one of the species under analysis. The promoter of the luxR gene of V. fischeri, for instance, has been shown to bind LexA and its own product (LuxR), which share the same binding site [70,71], and our analysis indicates that this cross-regulation is maintained in the closely related A. salmonicida. Similarly, we identify LexA-binding sites in the promoter of the aforementioned rstR gene in all the V. cholerae strains harboring the CTX prophage [37]. Close examination of these promoter regions in classical and El Tor biovars reveals that the LexA binding site mediating the regulation of rstR has been selectively maintained in spite of substantial changes to its promoter region [72]. In CTX phages of the classical biotype, the region encompassing the LexA-binding site contains several insertions with respect to El Tor biotypes, in which the combined RstR-LexA regulation has been analyzed in detail [37] (Additional file 1). These insertions, together with the specific sequence of the LexA-binding site, can in fact be used reliably to determine the CTX biotype of Vibrio sequences. The variability of the rstR promoter is in contrast with that observed in other SOS-regulated elements of V. cholerae. For instance, the promoter of the setR gene in the integrating-conjugative element SXT shows remarkable conservation of promoter elements and SetR-binding sites among strains (Additional file 2). Significant conservation is also observed within each biotype for the rstR promoter, suggesting that both variants of CTX and the SXT element have spread recently among V. cholerae strains. Our analysis also uncovers the recently reported regulation of Vibrio superintegrate genes [38,41] and the systematic regulation of the imuA-imuB-dnaE2 mutagenesis cassette in all the species in which it is present [49]. Interestingly, the presence of this mutagenesis cassette correlates neatly with the location of the superintegrate (found in the large chromosome in all the species harboring the cassette and in the small chromosome otherwise). This suggests that both traits are the product of a single reorganization event that took place before the split of V. cholerae from other Vibrionaceae, (such as Vibrio harveyi), a fact that is in consonance with the existing 16S rRNA phylogeny for this family [32].

**LexA binds predicted SOS genes in V. parahaemolyticus**

To validate the results of the *in silico* comparative genomics analysis, we performed electro-mobility shift assays (EMSA) of putatively regulated promoters of V. parahaemolyticus using the purified LexA protein of this organism. EMSA were performed on promoters of three canonical SOS genes (*lexA*, *recA* and *imuA*) for positive control, and on the promoters of genes flagged as candidate new members of the LexA regulon in the comparative genomics analysis: *topB*, *recG*, *mutH* and the two genes coding for proteins of unknown function (*unfA* [VP1264] and *unfB* [VP1428]). The results of the EMSA shown in Figure 3 confirm that the putative LexA-binding sites identified in the comparative genomics analysis are bound by LexA in V. parahaemolyticus, thus providing experimental validation for the *in silico* approach. Furthermore, competition assays establish that LexA-binding to these sites is specific (Additional file 3).
Hence, even though they do not constitute conclusive proof of SOS regulation, these results provide direct evidence of LexA-binding to the promoter region of five genes that have never been described previously as members of the SOS response. The apparent LexA regulation of recG makes intuitive sense. Even though it is not LexA regulated in E. coli, the product of recG is involved in the generation of Holliday junctions after replication fork stall and is known to interact with other members of the DNA recombination repair system [73]. The same logic applies to topB, which encodes a DNA topoisomerase III that has been shown in vitro to interact in E. coli with the products of ssb and recQ in the resolution of converging replication forks [74].

The putative SOS regulation of mutH, which encodes the DNA mismatch repair (MMR) endonuclease MutH, is more unexpected than that of recG or topB because no members of the MMR system have been reported to be SOS regulated previously. However, increased concentrations of MutH have been shown to lead to a predominance of MMR over very short patch repair (VSP), mediated by the Vsr endonuclease [75], and prevalence of MMR over VSP may be desirable to minimize the mutagenic effects of VSP. Finally, the V. parahaemolyticus EMSA confirm binding of LexA to the promoters of two genes coding for proteins of unknown function (VP1264 and VP1428). Homologs of the VP1264 gene, here termed unfA, have been annotated as members of the superfamily II of DNA and RNA helicases, for which recQ is the most well-known representative in bacteria. Hence, binding of LexA to the unfA promoter in the Vibrionaceae might be linked to the regulation of topB reported above and is likely to be involved in DNA repair processes. The gene VP1428 (unfB), however, has no known homologs outside the Vibrionaceae and its product is consistently annotated as hypothetical in all the Vibrio species analyzed here. Importantly, unfB appears to be present and associated with a putative LexA-binding site only in those Vibrio species that are known to be human pathogens. Given the well-founded relationship of the SOS response with dissemination of antibiotic resistance and pathogenicity in the Vibrio genus, it seems reasonable to postulate that the unfB gene product may be involved in such processes in Vibrio species.

The SOS regulon of Proteobacteria shares a small set of genes

Multiple chromosome genomes have been described and appear to have evolved independently in at least five different bacterial clades [31,76-79]. Besides the γ-Proteobacteria, to which the Vibrionaceae belong, complete genome sequences composed of multiple chromosomes are available for the α-Proteobacteria (Rhizobiaceae, Brucellaceae and Rhodobacteraceae families), the β-Proteobacteria (Burkholderiaceae and Comamonadaceae families), the Chloroflexi (Sphaerobacteraceae), the Deinococci (Deinococcaceae) and the Spirochaetes (Leptospiraceae). Having validated the comparative genomics approach in the Vibrionaceae, we decided to extend the analysis to other phylogenetic groups that present genomes with multiple chromosomes in order to analyze the adaptation of a complex genetic network, like the SOS response, to such genomic environments. The lack of a well defined LexA-binding motif and/or more than one complete genome sequence within a given phylogenetic group restricted our analysis to the β-Proteobacteria, to which the here-validated LexA-binding motif
can be applied [43] and to the α-Proteobacteria, in which a suitable LexA-binding motif has already been experimentally validated for comparative genomic approaches [64].

The results of the comparative genomics analysis for α- and β-Proteobacteria are presented, respectively, in Figure 4 and Figure 5. These support previous results reporting significant variation in the composition of the SOS system across bacterial groups [64,66]. In particular, the results on α- and β-Proteobacteria reveal a conserved core for the SOS regulon that comprises only the lexA and recA genes, an inducible TLS polymerase (dinB and/or polB), the NER excinuclease subunit A (uvrA) and the mutagenesis cassette imuA1-imuB-dnaE2.

Beyond this small core, the three phylogenetic groups analyzed here present numerous differences and some relevant similarities. A feature common to α- and β-Proteobacteria is the lack of LexA regulation of the recN gene, which is heavily regulated (up to three LexA-binding sites) in E. coli and in the Vibrionaceae, and which had been formerly identified as a key component of the SOS response [43]. On the other hand, the recG gene of some α-Proteobacteria appears to be regulated by LexA, suggesting that the presence of LexA-binding sites upstream of recG reported in the Vibrionaceae might be due to an ancestral regulation of this gene. The same reasoning can be applied in the case of the ruvCAB operon, the promoter of which harbors putative LexA-

---

**Figure 4** Tabulated description of the predicted LexA regulon of α-Proteobacteria species with multiple chromosome genomes. Colors indicate the presence and location of the gene and patterns denote presence (plain) or absence (patterned) of one or more LexA-binding sites in its promoter region. E. coli genes and their corresponding regulation are shown for comparative purposes. Eco, E. coli; Arad, Agrobacterium radiobacter; Atu, A. tumefaciens; Avi, Agrobacterium vitis; Bab, Brucella abortus; Bca, Brucella canis; Bme, Brucella melitensis; Bov, Brucella ovis; Bsu, Brucella suis; Oant, Ochrobactrum anthropi; Pden, Paracoccus denitrificans; Rsph, Rhodobacter sphaeroides; Sme, S. meliloti.
binding sites in the α-Proteobacteria in spite of substantial genomic rearrangements. In a similar vein, the identification of putative LexA-binding sites in the promoter of β-Proteobacteria Helicase c2 coding genes is congruent with the apparent regulation of unfA and recG in the Vibrionaceae.

Photoreactivation is under SOS regulation in the α- and β-Proteobacteria

In the α-Proteobacteria, the comparative genomics analysis identifies several genes not belonging to the shared SOS core that have already been reported as SOS regulated in some species with single chromosomes (parE in Sinorhizobium meliloti and tag, comM and dnaE in Caulobacter crescentus) [64,80]. Conversely, the analysis of the β-Proteobacteria suggests that two genes coding for hypothetical proteins might be LexA regulated in this class. One of them (hypoA) has no homologs outside β-Proteobacteria species with multiple chromosomes, whereas the other (hypoB) appears to be exclusive to the Burkholderia genus. Beyond the consistent lack of recN regulation, however, the predicted α- and β-Proteobacteria SOS networks share an additional trait that is not observed in the γ-Proteobacteria: the apparent regulation of a gene coding for a predicted photolyase (splB). The splB gene identified here codes for a protein homologous to the spore photoproduct lyase of B. subtilis, which is known to address DNA lesions generated by UV radiation in B. subtilis spores [81]. The fact that the homologs of this gene present LexA-binding sites in two groups where LexA recognizes markedly divergent LexA-binding motifs suggests that the splB gene is a functional member of the SOS response in both groups. This finding is remarkable because it is the first evidence of a photoreactivation enzyme regulated by the SOS response. Together with the predicted regulation of mutH in V. parahaemolyticus, it extends the conventional reach of this transcriptional response to include the direct regulation of MMR and photoreactivation components.

The apparent regulation of splB by the SOS response also points to an adaptive component in the specialization of this transcriptional network in different bacterial groups. As opposed to the Vibrionaceae, α- and β-Proteobacteria are mainly soil bacteria [82]. Soil habitats can lead to bacteria being exposed to high levels of UV

---

**Figure 5 Tabulated description of the predicted LexA regulon of β-Proteobacteria species with multiple chromosome genomes.** Colors indicate the presence and location of the gene and patterns denote presence (plain) or absence (patterned) of one or more LexA-binding sites in its promoter region. E. coli genes and their corresponding regulation are shown for comparative purposes. Eco, E. coli; B. 383, Burkholderia sp.; Bamb, Burkholderia ambifaria; Been, Burkholderia cenocepacia; Bglu, Burkholderia glumaeae; Bmal, Burkholderia mallei; Bmul, Burkholderia multivorans; Bphy, Burkholderia phytofirmans; Bpse, Burkholderia pseudomallei; Bthra, Burkholderia thailandensis; Bvie, Burkholderia vietnamiensis; Bxen, Burkholderia xenovorans; Reut, Raistonia eutrophica; Rmet, Raistonia metallidurans; Rpic, Raistonia picketti; Ctai, Cupnavidus taiwanensis; Vpar, Variovorax paradoxus.

| Gene Location | LexA-binding site | Present | Absent |
|---------------|-------------------|---------|--------|
| Chrom. 1      |                   |         |        |
| Chrom. 2      |                   |         |        |
| Chrom. 3      |                   |         |        |
| Plasmid       |                   |         |        |

**Legend:**
- **LexA** LexA repressor
- **recN** Recombination A
- **yebG** DNA repair protein
- **yigN** DNA recov RecC
- **dinB** DNA polymerase IV
- **polB** DNA polymerase II
- **uvrD** DNA helicase II
- **imuDC** DNA polymerase V
- **ruvc** Holliday recombination
- **ruvA** Holliday DNA helicase
- **imuA 1** Mutagenesis cassette
- **imuA 2** Mutagenesis cassette
- **imuA 3** Mutagenesis cassette
- **splB** Photolyase
- **Helicase c2** DNA helicase
- **hypoA** Unknown function A
- **hypoB** Unknown function B

---
light in combination with dehydration. This condition is encountered often by B. subtilis spores, and the ensuing DNA lesions are repaired by SlpB [81]. Hence, it seems plausible that α- and β-Proteobacteria might benefit from regulation of photoreactivation processes by the SOS response. The case of the splB gene thus provides a window into the process of adaptation of this transcriptional network to different ecological niches. More generally, the three LexA regulons analyzed in this work show evidence of radial divergence from a small shared core, and this divergent process has endowed each group with a different set of specialized SOS genes. The cooption of the SOS response to regulate dissemination of antibiotic resistance, pathogenicity and virulence determinants and its enhanced repertoire of repair genes in the Vibrionaceae has already been discussed. Presumably, the loss of regulation for some repair genes (e.g. ruvCAB) in the β-Proteobacteria and the regulation of DNA glycosilases (tag) and competence genes (comM) follows a similar, if not so obvious, adaptive rationale.

The SOS regulon core is contained within the larger chromosome

A main result of the comparative genomics analysis over several phylogenetic groups containing species with multiple chromosome genomes is the observation that the core genes of the SOS response are found predominantly in the large chromosome. Genomes with multiple chromosomes may originate through different mechanisms (e.g. chromosome duplication or partition), but the prevailing view on the evolution of multiple chromosomes in bacteria points to a plasmid origin of the smaller chromosomes [83]. This is supported by the presence of plasmid-like origins of replication in several small chromosomes [83], and by the greater proportion of housekeeping genes, synten and gene conservation in the large chromosomes [84,85]. Few comparative genomics studies have analyzed regulatory networks in species with multiple chromosomes and none of them has done so in a systematic way. Nonetheless, the scant data available suggest that genes belonging to primary regulatory networks, like those involved in anaerobic respiration, tend to reside in the large chromosome [86], whereas the genes composing more specialized regulons (like the KdgR regulon involved in pectin catabolism) are located mainly in the small chromosomes [57].

The comparative genomics analysis conducted here is thus the first systematic study of the distribution of a regulatory network in genomes with multiple chromosomes. The near-universality of the SOS response in the Bacteria domain and the well-documented housekeeping role of many of its members (e.g. recA) [23,87] establish it as an essential component of the genetic makeup of most bacteria. Hence, our analysis provides sound support to the hypothesis that the genomic distribution of primary regulons in species with multiple chromosomes is predominantly skewed towards the large chromosome. On the other hand, our study reveals also that family-specific additions to the SOS network, as well as duplications of core SOS genes, are most frequently located in small chromosomes and plasmids. The genetic content of small chromosomes has been shown recently to evolve faster, putting forward the notion that these secondary chromosomes act as evolutionary test beds subject to reduced purifying selection [88]. Here we build up on this idea to suggest that small chromosomes and plasmids might provide ideal grounds for the rewiring of transcriptional networks and, specifically, for systems like the SOS response that must strike a delicate balance between regulation of essential genes and rapid adaptability to ecological changes.

Conclusions

This work provides for the first time a detailed description of the SOS regulatory network in a bacterial family of clinical importance. This analysis identifies and experimentally validates new genes bound by LexA in the Vibrionaceae and supports the previously reported link between this stress response, the dissemination of antibiotic resistance and the adaptation of Vibrio species to human pathogenicity. The mapping of the SOS response in other bacterial groups with multiple chromosome genomes reveals a pattern of recent adaptation to specific environmental niches. Furthermore, it shows that the key elements of this response are consistently located in the large chromosome and it suggests smaller chromosomes and plasmids may serve as test beds for the rewiring of transcriptional networks.

Methods

Genome and LexA-binding data

Bacterial species with completely sequenced genomes composed of multiple chromosomes were identified through filtered queries on the NCBI Entrez system [89]. Complete chromosome and plasmid sequences for species with multiple chromosome genomes were downloaded from the RefSeq database in GenBank (.gbk) format [65]. Collections of binding sites for the γ/β- and the α-Proteobacteria were derived from experimental data from E. coli and several α-Proteobacteria species as described previously [43,64],

Comparative genomics analyses

Comparative genomics analyses were carried out using xFITOM, a program for binding site search in genomic sequences [59,90]. All chromosome sequences were searched for putative LexA-binding sites using the Rj index [91] and a motif-normalized threshold as reported.
The Electrophoresis mobility shift assays (EMSA) searches in the three phylogenetic groups analyzed in able LexA-binding site within a given family if candidate BLASTP [93]. A gene was considered to possess a reli-
point and their score was larger than the average score located within -300 to +50 bp of a gene translation start species were considered candidate sites if they were previously [92]. Putative LexA-sites identified in a single three different species. The detailed results for the searches in the three phylogenetic groups analyzed in this work can be found in Additional file 4.

Electrophoresis mobility shift assays (EMSA)
The *V. parahaemolyticus* lexA gene was amplified using suitable primers (Additional file 5) and cloned into a pET15b vector (Additional file 6). The *V. parahaemolyticus* LexA protein was over-expressed and purified as described previously [41]. DNA probes were constructed using two complementary 61 bp synthetic oligonucleo-
tides (Additional file 6). EMSA experiments were performed as reported previously [49], using 40 nM of *V. parahaemolyticus* LexA and 20 ng of each DIG-marked DNA probe in the binding mixture. For EMSA competi-
tive assays (Additional file 1), 200 fold of either specific or non-specific non-labeled DNA was added to the binding mixture. In all cases samples were loaded in 6% non-denaturing Tris-glycine poly-acrylamide gel. Digoxi-
genin-labeled DNA-protein complexes were detected following the manufacturer’s protocol (Roche).

Additional material

**Additional file 1: Analysis of rstr promoter regions**

(A) Comparison of the rstr-rstr intergenic region of *V. cholerae* O395 classical O1 serotype and *V. cholerae* O1 biovar El Tor strain N16961. Predicted (O395) and known (N16961) promoter elements (-35 and -10) are highlighted in grey. Known (N16961) [72] and predicted (O395) half-site operator sequences are highlighted in orange. Known (N16961) and predicted (O395) LexA-binding sites are shown in bold blue and underlined. RstR operators [36] are indicated by green boxes. (B) Alignment of 9 representative *V. cholerae* rstr-rstr intergenic regions. The LexA-binding site is shown in a blue box. The alignment shows a clear-cut distinction between El Tor and classical biotypes, but remarkably little divergence within each group, consistent with a recent dissemination of these phages. Sequence alignment was carried out using CLUSTALW [94] and default parameters. Adobe Portable Document File.

**Additional file 2: Analysis of setR promoter regions**

Alignment of the setR promoter region in different Vibrio cholerae strains. The SetR-binding sites [95] are highlighted in yellow. The putative promoter elements (-10 and -35) are bolded in blue. The translation start site for setR is bolded in red. Sequence alignment was carried out using CLUSTALW [94] and default parameters. Adobe Portable Document File.

**Additional file 3: EMSA competition experiments**

The lanes show, respectively, the standard EMSA using *V. parahaemolyticus* LexA and LexA promoter, the competition assay adding 200-fold excess of unlabelled lexA promoter, and the competition assay adding 200-fold excess of unlabelled non-specific DNA. JPEG image.

**References**

1. Radman M: Phenomenology of an inducible mutagenic DNA repair pathway in *Escherichia coli*: SOS repair hypothesis. In *Molecular and environmental aspects of mutagenesis*. Edited by: Prokash L, Sherman F, Miller M, Lawrence C, Tabor HW; Charles C. Thomas Publisher, Springfield, IL: 1974:128-142.

2. Walker GC: The SOS response of *Escherichia coli*, In *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*. Volume 1. Edited by: Neidhart FC, Ingrum JL, Low KB, Magasanik B, Schaechter M, Umbarger HE. Washington, D.C.: American Society for Microbiology, 1987.

3. Little JW, Mount DW: The SOS regulatory system of *Escherichia coli*. Cell 1982, 29:11-22.

4. Wettman KE, Mount DW: Nucleotide sequence binding specificity of the LexA repressor of *Escherichia coli* K-12. Journal of bacteriology 1985, 163:376-384.

5. Fernandez De Henestrosa AR, Ogi T, Aoyagi S, Chafin D, Hayes JJ, Ohmori H, Woodgate R: Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. Molecular microbiology 2000, 35:1560-1572.

6. Little JW: Control of the SOS regulatory system by the level of RecA protease. Biochime 1982, 64:585-589.

7. Anderson DG, Kovalczykowski SC: Reconstitution of an SOS response pathway: derepression of transcription in response to DNA breaks. Cell 1998, 95:975-979.
8. Sassanfar M, Roberts JW: Nature of the SOS-inducing signal in Escherichia coli. The involvement of DNA replication. *Journal of molecular biology* 1990, 212:79-96.

9. Little JW: Mechanism of specific LexA cleavage: autodigestion and the role of RecA coprotease. *Biochimie* 1991, 73:41-421.

10. Crowley DJ, Courcelle J: Answering the Call: Coping with DNA Damage at the Most Inopportune Time. *J Bacteriol* 2002, 26:65-74.

11. Little JW, Mount DW, Yanisch-Perron CR: Purified LexA protein is a repressor of the recA and lexA genes. *Proceedings of the National Academy of Sciences of the United States of America* 1981, 78:4190-4203.

12. Costa de Oliveira R, Laval J, Bontems S: Induction of SOS and adaptive responses by alkalizing agents in *Escherichia coli* mutants deficient in 3'-methyladenine-DNA glycosylase activities. *Mutation research* 1987, 183:11-20.

13. Jarosz DF, Beuning PJ, Cohen SE, Walker GC: Y-family DNA polymerases in *Escherichia coli*. *Trends in microbiology* 2007, 15:70-77.

14. Bjedov I, Tenaillon O, Gerard B, Souza V, Denamur E, Radman M, Taddei F, Sanchez-Alberola A: Eons of distress: an evolutionary perspective of the ultraviolet induction of prophage lambda. *Science (New York, NY)* 2006, 310:1404-1409.

15. Erill I, Campoy S, Mazon G, Barbie J: Dispersal and regulation of an adaptive mutagenesis cassette in the bacteria domain. *Nucleic acids research* 2006, 34:67-77.

16. Herman I, Luria SE: Transduction studies on the role of a rec+ gene in the ultraviolet induction of prophage lambda. *Journal of molecular biology* 1967, 23:117-133.

17. Muhlendorf I, Haekker J, Keshch GT, Acheson DW, Tschape H, Kake AV, Ritter A, Olschlager T, Donohue-Rolfe A: Regulation of the Shiga-like toxin II operon in *Escherichia coli*. *Infection and immunity* 1996, 64:495-502.

18. Ubeda C, Marques E, Tormo MA, Campon S, Lasa J, Barbie J, Novick RP, Penades JR: SaPI operon I is required for SaPI packaging and is controlled by LexA. *Molecular microbiology* 2007, 65:41-50.

19. Miller C, Thomsen LE, Gaggero C, Mosseri R, Radman M, Taddei F, Marc I: Stress-induced mutagenesis in bacteria. *Science (New York, NY)* 2003, 300:1404-1409.

20. Erill I, Campoy S, Mazon G, Barbie J: Dispersal and regulation of an adaptive mutagenesis cassette in the bacteria domain. *Nucleic acids research* 2006, 34:67-77.

21. Muhlendorf I, Haekker J, Keshch GT, Acheson DW, Tschape H, Kake AV, Ritter A, Olschlager T, Donohue-Rolfe A: Regulation of the Shiga-like toxin II operon in *Escherichia coli*. *Infection and immunity* 1996, 64:495-502.

22. Ubeda C, Marques E, Tormo MA, Campon S, Lasa J, Barbie J, Novick RP, Penades JR: SaPI operon I is required for SaPI packaging and is controlled by LexA. *Molecular microbiology* 2007, 65:41-50.

23. Erill I, Escobero M, Campoy S, Barbie J: In silico analysis reveals substantial variability in the gene contents of the gamma proteobacteria *LexA-regulon*. *Bioinformatics* (Oxford, England) 2003, 19:2225-2236.

24. Campoy S, Mazon G, Fernandez de Henestrosa AR, Llagostera M, Giner GR, Barbie J: A new regulatory DNA motif of the gamma subclass Proteobacteria: Identification of the LexA protein binding site of the plant pathogen *Xylella fastidiosa*. *Microbiology (Reading, England)* 2002, 148:3583-3597.

25. Cheo DL, Bayles KW, Yabum RE: Cloning and characterization of DNA damage-inducible promoter regions from *Bacillus subtilis*. *Journal of bacteriology* 1991, 173:1696-1703.

26. Mavahdezadeh F, Colston MJ, Davis EO: Characterization of *Mycobacterium tuberculosis LexA*: recognition of a Cheo (Bacillus-type SOS) box. *Microbiology (Reading, England)* 1997, 143(Pt 4):929-936.

27. Tapis A, Barbie J: Mutational analysis of the *Rhizobium etli recA* operator. *Journal of bacteriology* 1998, 180:6325-6331.

28. Campoy S, Salvador N, Cortes P, Erill I, Barbie J: Expression of canonical SOS genes is not under LexA repression in *B. lindbyi* *bacteriovorus*. *Journal of bacteriology* 2005, 187:5367-5373.

29. Abella M, Erill I, Jara M, Mazon G, Campoy S, Barbie J: Widespread distribution of a lexA-regulated DNA damage-inducible multiple gene cassette in the Proteobacteria phylum. *Molecular microbiology* 2004, 54:212-222.

30. Mazon G, Campoy S, Erill I, Barbie J: Identification of the *Acidobacterium capsulatum LexA* box reveals a lateral acquisition of the Alphaproteobacteria *lexA* gene. *Microbiology* (Reading, England) 2006, 152:1109-1112.

31. McGuire AM, Hughes JD, Church GM: Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. *Genome research* 2000, 10:444-757.

32. Tan K, Moreno-Hagelsieb G, Collado-Vides J, Stormo GD: A comparative genomics approach to prediction of new members of regulons. *Genome research* 2001, 11:566-584.

33. Clifton PF, Hillier LW, Fulton L, Graves T, Miner T, Gish WR, Waterston RH, Johnston M: Surveying *Saccharomyces* genomes to identify functional motifs. *Genome research* 2001, 11:1175-1186.

34. Siddharthan R, Sigga ED, van Nimwegen E: Phyloglobin: A Gibbs sampling motif finder that incorporates phylogeny. *PLoS computational biology* 2005, 1:e67.

35. Jensen ST, Shen L, Liu JS: Combining phylogenetic motif discovery and motif clustering to predict co-regulated genes. *Bioinformatics* (Oxford, England) 2005, 21:3832-3838.

36. Sinha S, Blanchette M, Tompa M: PhyME: a probabilistic algorithm for finding motifs in sets of orthologous sequences. *MCB bioinformatics* 2004, 5:170.

37. Rodionov DA, Gelfand MS, Hugouvieux-Cotte-Pattat N: Comparative genomics of the *KdgR* regulon in *Erwinia chrysanthemi* 3973 and other
gamma-proteobacteria. *Microbiol. (Reading, England)* 2004, **150**:3571-3590.

58. Panina EM, Mironov AA, Gelfand MS: Comparative genomics of bacterial zinc regulations: enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins. *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**:9912-9917.

59. Erll I, O'Neill MC: A reexamination of information-theory-based methods for DNA-binding site identification. *BMC bioinformatics* 2009, **10**:57.

60. Oaida R, Zašlavsky E, Singh M: Comparative analysis of methods for representing and searching for transcription factor binding sites. *Bioinformatics (Oxford, England)* 2004, **20**:3516-3525.

61. Mazon G, Lucena JM, Campoy S, Fernandez de Henestrosa AR, Candau P, Barbe J: LexA-binding sequences in Gram-positive and cyanobacteria are closely related. *Mel Genet Genomes* 2004, **271**:40-49.

62. Garriga X, Calero S, Barbe J: Nucleotide sequence analysis and comparison of the lexA genes from *Salmonella typhimurium, Erwinia carotovora, Pseudomonas aeruginosa* and *Pseudomonas putida*. *Mel Genet Genomes* 1992, **236**:125-134.

63. Tapia A, Campoy S, Barbe J: Analysis of the expression of the *Rhodobacter sphaeroides* lexA gene. *Mel Genol Genomes* 2000, **263**:957-965.

64. Erll I, Jara M, Salvador N, Escribano M, Campoy S, Barbe J: Differences in LexA regulon structure among Proteobacteria through in vivo assisted comparative genomics. *Nucleic acids research* 2004, **32**:6617-6626.

65. Prutt KD, Tatsioua T, Maglott DR: NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research* 2005, **33**:D501-504.

66. Au N, Kuester-Schoeck E, Canny SP, Chachu K, Colavito SA, Fuller SN, Groban ES, Hensley LA, et al: Genetic composition of the Bacillus subtilis SOS system. *Journal of bacteriology* 2005, **187**:665-6666.

67. Perrino FW, Rein DC, Robst AM, Meyer RR: The relative rate of synthesis and levels of single-stranded DNA binding protein during induction of SOS repair in Escherichia coli. *Mel Genol Genomes* 1987, **209**:612-614.

68. Brandriss JA, Bosch D, de Ruiter M, van de Putte P: Analysis of the regulatory region of the ssb gene of *Escherichia coli*. *Nucleic acids research* 1983, **13**:5005-5109.

69. Baharoglu Z, Petranovic M, Flores MJ, Michel B: RuvAB is essential for replication forks reversal in certain replication mutants. The *EMBO* journal 2006, **25**:596-604.

70. Shadel GS, TOB, JHD: Control of the lux regulon of *Vibrio fischeri*. *Journal of Bioluminescence and Chemiluminescence* 1990, **5**:99-106.

71. Ulltuz S: Involvement of Host Factors in Regulating the lux System of *Vibrio fischeri*. In: *Microbial ecology and infectious disease*. Edited by: Rosenberg E. American Society for Microbiology. 1999.

72. Kimsey HH, Waldor MK: The CTXpHi repressor RsrR binds DNA cooperatively to form tetrameric repressor-operator complexes. *The Journal of biological chemistry* 2004, **279**:2650-2657.

73. McGlynn P, Lloyd RC: Rescue of stalled replication forks by RecG: simultaneous translocation on the leading and lagging strand templates supports an active DNA unwinding model of fork reversal and Holliday junction formation. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**:8227-8234.

74. Suski C, Marians KJ: Resolution of converging replication forks by RecQ and topoisomerase III. *Molecular cell* 2006, **30**:778-789.

75. Bhagwat AS, Lieb M: Cooperation and competition in mismatch repair: very short-patch repair and methyl-directed mismatch repair in Escherichia coli. *Molecular microbiology* 2002, **44**:1421-1428.

76. Barli C, Herrmann JL, Richaud C, Margarita D, Geiss IS: Scattering of the rRNA genes on the physical map of the circular chromosome of *Leptospira interrogans* serovar icterohaemorrhagiae. *Journal of bacteriology* 1992, **174**:756-7571.

77. Rodley PD, Rumlins U, Tummler B: A physical genome map of the *Burkholderia cepacia* type strain. *Microbiol. (Reading, England)* 1995, **151**:17-67.

78. White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, Dodson RI, Haft DH, Gwinn ML, Nelson WC, Richardson DL, et al: Genome sequence of the radioreistant bacterium *Deinococcus radiodurans* R1. *Science (New York, NY)* 1999, **286**:1571-1577.

79. Past A, Labbati K, PKuli R, Nolan M, Glavina Del Rio T, Tice H, Cheng JF, Lucas S, Chen F, Copeland A, et al: Complete genome sequence of *Sphaerobacter thermophilus* strain (S 6022). *Standards in genomic sciences* 2010, **2**:49-56.

80. Paes da Rocha R, Paquola AC, do V Marques M, Menck CF, Galhardo RS: Characterization of the SOS regulon of *Caulobacter crescentus*. *Journal of bacteriology* 2008, **190**:2129-2138.

81. Chandor A, Beretua O, Douki T, Gasparutto D, Sanakis Y, Ollagnier-de-Choudens S, Atta M, Fontecave M: Dinucleotide spore photoproduc, a minimal substrate of the DNA repair spore photoprodut lyase enzyme from *Bacillus subtilis*. *The Journal of biological chemistry* 2006, **281**:26922-26931.

82. Youssif NH, Elshahed MS: Diversity rankings among bacterial lineages in soil. *The ISME journal* 2009, **3**:305-313.

83. Esan ES, Fogle MA, Waldor MK: Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. *Molecular microbiology* 2005, **56**:1129-1138.

84. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, et al: DNA sequence of both chromosomes of the cholaera pathogen *Vibrio cholerae*. 2000, **406**:477-483.

85. Holden MT, Tittball RW, Peacock SJ, Cerdono-Tarraga AM, Atkins T, Crossman LC, Pitt T, Churker M, Mungall K, Bentley SD, et al: Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**:14240-14245.

86. Ravcheva DA, Gerisimova-AV, Mironov AA, Gelfand MS: Comparative genomic analysis of regulation of anaerobic respiration in ten genomes from three families of gamma-proteobacteria (Enterobacteriaceae, Pasteurellaceae, Vibrionaceae). *BMC genetics* 2007, **8**:54.

87. Lloyd AT, Sharp PM: Evolution of the recombination and the molecular phylogeny of bacteria. *J Mol Evol* 1993, **37**:399-407.

88. Cooper VS, Vooh SH, Wecklage SC, Hatcher PJ: Why genes evolve faster on secondary chromosomes in bacteria. *Plos computational biology* 2010, **6**:e1000732.

89. Entrez cross-database search. [http://www.ncbi.nlm.nih.gov/sites/gquery].

90. Bhargava N, Erill I: xFITEM: a generic GUI tool to search for transcription factor binding sites. *Bioinformatics* 2010, **4**:49-50.

91. Schneider TD: Information Content of Individual Genetic Sequences. *Journal of Theoretical Biology* 1997, **189**:427-441.

92. Abella M, Campoy S, Erll I, Rojo F, Barbe J: Cohabitation of two different lexA regulons in *Pseudomonas putida*. *Journal of bacteriology* 2007, **189**:8853-8862.

93. Fuchs CA, Rocap G: Whole-genome reciprocal BLAST analysis reveals that planctomycetes do not share an unusually large number of genes with *Eukarya* and *Archaea*. *Applied and environmental microbiology* 2006, **72**:6841-6844.

94. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 1994, **22**:4673-4680.

95. Beaber JW, Waldor MK: Identification of operators and promoters that control SXT conjugal transfer. *Journal of bacteriology* 2004, **186**:5945-5949.