BMC Genomics

Comparative analyses imply that the enigmatic sigma factor 54 is a central controller of the bacterial exterior

Christof Francke[1,2,4,6*], Tom Groot Kormelink[1,2,3,6], Yanick Hagemeijer[6], Lex Overmars[1,3,6], Vincent Sluijter[6], Roy Moezelaar[1,5] and Roland J Siezen[1,2,4,6]

Abstract

Background: Sigma-54 is a central regulator in many pathogenic bacteria and has been linked to a multitude of cellular processes like nitrogen assimilation and important functional traits such as motility, virulence, and biofilm formation. Until now it has remained obscure whether these phenomena and the control by Sigma-54 share an underlying theme.

Results: We have uncovered the commonality by performing a range of comparative genome analyses. A) The presence of Sigma-54 and its associated activators was determined for all sequenced prokaryotes. We observed a phylum-dependent distribution that is suggestive of an evolutionary relationship between Sigma-54 and lipopolysaccharide and flagellar biosynthesis. B) All Sigma-54 activators were identified and annotated. The relation with phosphotransfer-mediated signaling (TCS and PTS) and the transport and assimilation of carboxylates and nitrogen containing metabolites was substantiated. C) The function annotations, that were represented within the genomic context of all genes encoding Sigma-54, its activators and its promoters, were analyzed for intra-phylum representation and inter-phylum conservation. Promoters were localized using a straightforward scoring strategy that was formulated to identify similar motifs. We found clear highly-represented and conserved genetic associations with genes that concern the transport and biosynthesis of the metabolic intermediates of exopolysaccharides, flagella, lipids, lipopolysaccharides, lipoproteins and peptidoglycan.

Conclusion: Our analyses directly implicate Sigma-54 as a central player in the control over the processes that involve the physical interaction of an organism with its environment like in the colonization of a host (virulence) or the formation of biofilm.

Keywords: biofilm, enhancer binding protein, exopolysaccharide, lipopolysaccharide, nitrogen assimilation, motility, peptidoglycan

Background

Sigma factors specify bacterial transcription by binding to a characteristic promoter and thereby recruiting the associated RNA polymerase to that promoter. Ordinarily, the expression of genes/operons is controlled by the so-called 'housekeeping' sigma factor 70. However, most bacteria possess a larger repertoire of sigma factors of the Sigma-70 family, where each additional factor is associated with a specific programmed response [1]. For

instance, in *Escherichia coli* and related Gamma-proteobacteria the entry into stationary phase and the adaptation to starvation is associated with Sigma-S [2,3], whereas the response to heat shock and similar stresses is mediated by Sigma-32 (e.g. [4,5]). In *Bacillus subtilis*, sporulation is orchestrated by 5 sigma factors (Sigma-E, F, G, H and K) [6], whereas the general stress response is controlled by Sigma-B [7,8]. In many species, particular extracellular signals are translated into an appropriate response by ECF sigma factors [9].

There is one sigma factor that seemingly does not fit in this picture as it has been associated with a range of

* Correspondence: c.francke@cmbi.ru.nl
[1]TI Food and Nutrition, P.O.Box 557, 6700AN Wageningen, The Netherlands
Full list of author information is available at the end of the article

physiological phenomena instead of with a singular response. Sigma-54 (gene *rpoN* in *E.coli*, *sigL* in *B. subtilis*) constitutes an evolutionary separate protein family and is found widely distributed among the bacterial kingdom, although there are phyla that lack the protein [10,11]. It binds to a characteristic -24/-12 promoter [12-14] and absolutely requires the input of free energy (ATP) from an associated activator to initiate transcription [15,16] (see [17,18] for recent reviews on the mechanism). In most cases the activator binds to an enhancer element located upstream of the promoter and hence is referred to as Enhancer Binding Protein (EBP[54]). The EBP[54]s bind the DNA as inactive dimers, but upon reception of the appropriate signal they assemble into oligomeric rings [19,20], with hexamers constituting the oligomeric active state [21]. A large variety of EBP[54]s exists and although some species possess one, for instance *Chlamydia trachomatis* [22] and *Lactobacillus plantarum* [23], most species have more variants. *B. subtilis* and *E. coli* were reported to have five (see [24]) and twelve [25], respectively, and *Myxococcus xanthus* to have fifty-three [26]. However, many of the reported numbers need correction (as described later) because the previous analyses have included EBP[54] paralogs that have lost the interaction with Sigma-54, like TyrR [27] and DhaR [28] in *E. coli* and HupR in *Rhodobacter capsulatus* [29,30].

Historically, Sigma-54 has been linked to the regulation of nitrogen metabolism. The protein was discovered as a positive regulatory factor needed for the expression of enterobacterial glutamine synthetase [31], before it was recognized that the protein is actually a sigma factor [32]. However, it was soon after established that Sigma-54 mediated control of transcription is not only connected to nitrogen assimilation but to a wider range of cellular processes and physiology in the enterobacteria [25,33]. Since then, it was shown that its role also encompasses the regulation of for example: flagellar biosynthesis in *E. coli* [34]; carboxylate uptake, central metabolism and flagellar biosynthesis in *Geobacter sulfurreducens* [35]; phosphotransferase system (PTS)-mediated carbohydrate uptake in the Gram-positive species *Lactobacillus plantarum* [23] and *Listeria monocytogenes* [36]; and PTS-mediated regulation in Gram-positive as well as Gram-negative organisms [37,38]; osmotolerance in *Listeria* [39]; the utilization of compounds like gamma-aminobutyrate in *Bacillus* [40], and the less familiar biphenyl in *Ralstonia metallidurans* [41] and toluene, xylene (see [42]) and choline [43] in *Pseudomonas*; Type III secretion system mediated pathogenicity in *Pseudomonas syringae* [44] and Type VI secretion system mediated toxin secretion in e.g. *Aeromonas* and *Marinomonas* [45]; the adaptation to cold shock in *B. subtilis* [24]; the control of Sigma-S

[46], lipoprotein biosynthesis and virulence [47] in *Borrelia burgdorferi*; acid resistance of pathogenic *E. coli* O157 [48]; biofilm formation by *Burkholderia* [49]; and motility, biofilm formation, luminescence, and colonization in *Vibrio fischeri* [50,51]. The above plethora of associations has up to now obscured the definition of a general underlying functional theme that adds to the accepted associations with nitrogen metabolism and flagellar biosynthesis.

Several comparative studies have been performed for Sigma-54 and EBP[54]-mediated regulation [10,15,16,52], but no unifying biological theme was identified. An in-depth comparative analysis was made for *E. coli* by [25]. These authors concluded that nitrogen assimilation was one of the main processes connecting the Sigma-54 regulon. Besides, they found that a substantial fraction of the associated functions was seemingly unrelated. Some additional associations were proposed on basis of a comparative analysis on *Pseudomonas putida*, including links to carbon metabolism and flagellar biosynthesis [53]. Since the last comprehensive comparative study in 2003 a considerable number of genomes has been sequenced, allowing us to make a new overview of the presence of Sigma-54 and the EBP-activators. Surprisingly, we found a clear-cut connection between the presence of the system and characteristic morphological features. To enhance the identification of true EBP[54] activators and Sigma-54 promoters, we have tested and employed a straightforward motif search algorithm that directly relates to sequence similarity. Redefinition of the -24/-12 promoter and the similar motif search (SMS) approach allowed for the reliable identification of promoter sites in all species. Finally, we have analyzed the function annotations that were highly represented (intra-phylum) and conserved (inter-phylum) within the genomic context of all genes encoding Sigma-54, its activators and its promoters, to identify common functional traits.

Conserved genome context, i.e. synteny, is a strong indicator of a functional relationship between genes [54,55] and it is therefore being used broadly to guide function prediction. In principle, the fact that encoded functions that show a conserved genomic proximity are mostly related does not only hold for genes, but by necessity extends in the direction of genetic (regulatory) elements [56], and thereby also in the direction of associated regulators (see e.g. [57]) and their (in)activating signals [58]. As a consequence, a comparative analysis of the conserved genome context of regulators and regulatory elements should yield clues regarding the particular associated stimuli and responses. Although regulatory routes can vary between species much more than metabolic pathways, the functional associations at a higher hierarchical level (i.e. in terms of process, response and/

or physiology) are far less variable. For instance, the bacterial PTS mediates the transport and phosphorylation of carbohydrates by means of phosphoenolpyruvate via the same phosphotransfer mechanism in all species and, at a higher hierarchical level, the system controls the same processes like catabolite repression and chemotaxis [37,59]. Nevertheless, the precise regulatory interactions of the PTS and the intracellular signals that connect the organism's physiological state to the metabolic level differ significantly between groups of species (i.e. catabolite repression involves EIIA^{Glc} and cAMP in *E. coli*, whereas it involves HPr and Fructose-1,6-bisphosphate in *B. subtilis*). The above implies that underlying functional themes that can not be discovered directly, for instance by studying conserved gene-associations of a particular regulator, may be discovered by mapping the associated functions at a higher hierarchical level (like pathways).

Absolute conservation will be relatively rare because of the earlier noted variability in the specific regulatory associations. To take such variability into account, we included in our analysis those functional associations that are highly represented within a phylum/class but are at the same time evolutionary conserved, that is present within several phyla/classes. Associations that fulfill this criterion can be viewed as cross-phylum (or cross-class) conserved function tendencies. By mapping of the conserved annotations present in the genetic context of the genes encoding Sigma-54, its EBP^{54}-activators and its promoters, we discovered that there is indeed a common functional theme related to Sigma-54-mediated regulation, namely, the control of the transport and biosynthesis of the molecules that constitute the bacterial exterior, which encompass the extracellular polysaccharides (EPS), flagella, lipopolysaccharides (LPS), lipoproteins and the building blocks of the peptidoglycan cell wall.

Results

Taxonomic diversity of Sigma-54 and its Enhancers

Completely sequenced and published prokaryotic genomes listed in the GOLD database [60] were searched at the protein level for homologs of Sigma-54 and the related enhancer-binding proteins (EBP^{54}s). Experimentally characterized proteins were used to seed the search (see methods). In the case of Sigma-54, the sequences of the *E. coli* and *B. subtilis* proteins sufficed to recover all orthologs using a low cutoff value (1e^{-20}) (hits given in additional file 1). In the case of the far more diverse EBP^{54}s, a list of 34 experimentally verified proteins from diverse species was compiled and it required extensive filtering to reduce the initial list of putative homologs (see methods and below). The final list is given in additional file 2 and corresponds well with earlier analyses performed at a smaller scale [16,52].

The collection of sequences exhibited a near perfect match between the presence in a genome of Sigma-54 and its EBP^{54}-activators. A well-known exception was formed by the Neisseria strains that have a gene encoding Sigma-54 but lack an obvious candidate EBP^{54}. However, the *Neisserial* Sigma-54 protein is inactive because the two helices that interact with the -13 promoter region [61] are missing in this protein [62]. The other exceptions were *Borrelia hermsii* DAH, *Borrelia turicatae* 91E135, *Fusobacterium nucleatum nucleatum* ATCC 25586 and *Jannaschia* sp. CCS1, which possess a putative EBP^{54} but lack a Sigma-54 ortholog. As the initial search was performed using the translated protein sequences we also searched the nucleotide sequence directly and found that the chromosome of *B. hermsii* carries an intact *rpoN* gene that was obviously missed in the original annotation procedure. However, we could not identify *rpoN* in *B. turicatae*. A rationale for the absence of Sigma-54 in *Fusobacterium nucleatum nucleatum* ATCC 25586 and *Jannaschia* sp. CCS1 will be given in the next section.

Sigma-54 and EBP^{54}s were found in the majority of sequenced species (522 of 842 evaluated genomes) and in the majority of the evaluated phyla, as illustrated in Figure 1. Remarkably, their presence is almost absolutely conserved in species that are diderm (i.e. that have two cellular membranes [63]) and have an outer membrane that mainly consists of lipopolysaccharides (LPS) [64]. The diderm species in which Sigma-54 and its activators are absent represent mostly endosymbionts (> 90%), e.g. species from the Genus Bartonella, Buchnera, Ehrlichia, Francisella, Neisseria and Rickettsia, and *Elusimicrobium minutum* [65]. Furthermore, the proteins are also absent from all 29 sequenced species/strains of the phylum Cyanobacteria. For the monoderm phyla and for the diderm phyla that have different LPS or lack LPS altogether the picture is inverse, i.e. in most of the related species Sigma-54 and EBP^{54}s are absent. They were identified only in some Chloroflexi and Thermotogae. In contrast, they are present in many Firmicutes, that is in most species of the class Clostridia and in the sporulating species of the class Bacilli. In addition, they were found in Listeria, *Enterococcus faecalis* and in the closely related Lactobacilli *Lactobacillus plantarum*, *Pediococcus pentosaceus* and *Lactobacillus casei* (only strain BL23).

There was an overall correlation between the size of the genome and the presence of Sigma-54 as illustrated in Figure 2A. Most endosymbionts or facultative intracellular species have a small genome (< 1.5Mb) and lack Sigma-54, although some (i.e. Borrelia and Chlamydia) do have Sigma-54 and contain one activator. In addition, species of the phyla Actinobacteria and Cyanobacteria have much larger genomes (up to 9 Mb) but lack the Sigma-54 system. A small number of
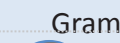
Figure 1 Taxonomic distribution of Sigma 54 and the associated Enhancer Binding Proteins. The phyla Proteobacteria and Firmicutes have been divided into the constituent classes. Between brackets, the number of genomes with Sigma-54 over the total number of sequenced genomes is listed followed by the average number of EBP[54]s with a discrete standard deviation. The corresponding data can be found in additional file 1. The ordering of the phyla is based on conserved gene arrangement comparisons [162], a concatenated alignment of 22 single-copy conserved genes [163] and the analysis of conserved indels [164]. Gemmatimonas was placed according to [165], Nitrospira according to [166] and Thermatogae according to [167]. On the right, the cell morphology in terms of number of membranes (monoderm vs. diderm), presence of LPS (from [64]) and nature of the cell wall peptidoglycan (Gram-, Gram+ or other) is given. The majority of phyla represent diderms, except for Tenericutes, Firmicutes and Actinobacteria. Chloroflexi are probably also monoderm [168] and some have been shown to have a thick cell wall and stain Gram positive [169]. *Deinococcus radiodurans* has a complex Gram + like cell wall that includes outer membrane-like structure and the cell wall and outer membrane can be shared by multiple cells [170]. *Dictyoglomus thermophilum* is diderm but can grow in bundles or spherical bodies which are surrounded by a common outer membrane [171]. Finally, the Thermotoga have an outer sheath-like envelope ('toga') and an atypical thin cell wall [172].

Figure 2 Distribution of genome size (A), the number of EBP[54]-activators (B), and motility (C), for species with (blue) and without (red-brown) Sigma-54. A) the analyzed species were binned according to genome size in bins of one Mbase, and divided in two groups that related to the presence or absence of Sigma-54. B) for every size-bin the fraction of genomes with a particular number of EBP-activators was determined and a height-plot was created. The grey dots indicate the data points. The contour was generated with Microsoft Office Excel 2007. C) Within every bin the fraction of motile species was determined for the genomes with Sigma-54 and without. The corresponding data can be found in additional file 1.

Proteobacteria, mostly species/strains from the orders Burkholderiales, Rhizobiales and Xanthomonadales (see also [10,66-68]), possess two variants of the sigma factor and the sequenced *Rhodobacter sphaeroides* strains even contain three or four variants [69,70]. In the case that Sigma-54 is present, there is a strong positive correlation between the number of encoded EBP[54]s and the size of the genome as might be expected (see Figure 2B). On top of that, there appeared to be a clear phylum/class dependency ranging from the Delta-proteobacteria with around 33 EBP[54]s on average to the Epsilon-proteobacteria, Chlamydia and Spirochetes with on average 1 or 2 activators (Figure 1). Importantly, a strong positive correlation between reported motility and the presence of Sigma-54 was observed and this correlation appeared predominantly independent of genome size (Figure 2C). The strong correlation fortifies the general view that one of the common functional themes of Sigma-54 mediated control is the association with the synthesis of flagella and or pili, an association that has been observed for many species (reviewed in [71]). Other clear correlations with

particular bacterial lifestyles (as represented in the GOLD database) were not observed.

Sequence-based identification and characterization of the EBP[54]s

The list of EBP[54]s that was obtained in a BLAST search using the Sigma-54 interaction/activator domain (PFAM: PF00158), was initially filtered using two criteria based on cut off value (see methods). In this way, 5494 potential EBP[54]s were identified. The list contained many false positives (~10%), which were mostly proteases and Mg^{2+} chelatases [72] as these are the closest relatives of the Sigma-54 interaction/activator domain [73]. To remove false positives, we analyzed the presence of the characteristic 'GAFTGA' amino acid sequence that is essential for the interaction between the activator and the sigma factor [74]. It has been established that single residue changes within the sequence element and especially within the central phenylalanine and threonine, reduce the transcriptional activity considerably and mostly abolish it [75,76]. Given the reported constraints, which are summarized in the

legend of Figure 3, 4850 true and 121 putative EBP[54]s were obtained (additional file 2). The 'GAFTGA' sequence within this set is highly conserved (see Figure 3A), and the importance of this conservation is corroborated by the substitutions/deletions that lead to 'nonfunctional' EBP[54] homologs, like in TyrR, DhaR of *E. coli* and HupR of *R. capsulatus*, which have been shown to be Sigma-70 dependent ([27,28] and [29,30], respectively). Similar changes were observed within the protein sequence of the VpsR regulator of polysaccharide

Figure 3 Sequence composition of the Sigma-54 Enhancer Binding Proteins. A) The 'GAFTGA' sequence logo of the 4970 putative functional Sigma-54 related EBPs. Data from literature and similarity in chemical structure were used to categorize the substitutions into those that relate to functional EBP[54]s, those that will probably relate to functional EBP[54]s, and those that will abolish the interaction with Sigma-54. The effect of amino acid substitutions on the EBPs capacity to activate Sigma-54 mediated transcription has been studied by [75,76]. Furthermore, some experimentally validated activators carry specific substitutions: G_1 is replaced by N in the only EBP[54] of *Paracoccus denitrificans* and *Ruegeria pomeroyi* (putative: ADEHS); A_2 is replaced by S in LevR of the Bacilli (putative: TGIVMC; inactive: DN); F_3 is replaced by Y in TouR of *Pseudomonas stutzeri* (other replacements inactive); T_4 is replaced by S in BkdR of *B. subtilis* and by E in PhhR of *Pseudomonas aeruginosa* (putative: D; other replacements inactive); G_5 is replaced by D in FlgR, the only EBP[54] of Campylobacter and other Epsilon-proteobacteria (putative; EAHNS); and A_6 is replaced by S in PrpR of *E. coli* (putative: TGIVMC; inactive DN). B) Schematic representation of the four basic architectures of functional EBP[54]s. The types were distinguished on basis of their domain organization: Ia) N-terminal signal recognition domain of the response regulator (RR) type, followed by the central activator domain and a C-terminal DNA-binding domain of the HTH_8 PFAM family; Ib) different N-terminal signal recognition domain(s), followed by the central activator domain and a C-terminal DNA-binding domain of the HTH_8 PFAM family; Ic) an activator domain, but lacking the signal recognition domain (e.g. PspF, HrpRS, LafK) or the DNA-binding domain (e.g. CtcC, FlgR) or both (FleT); and II) N-terminal DNA-binding domain of the NtrC family, the central domain, and four phosphorylatable domains related to the PTS.

production in *Vibrio cholerae* and the YplP regulator of the cold shock response in *B. subtilis*. These proteins were proposed to be Sigma-54 dependent ([77] and [24], respectively), but considering the lack of compelling experimental evidence and the clear deletion within the 'GAFTGA' sequence (see additional file 3) they are probably not. We found that around 110 of the recovered homologs showed such minor deviations and within this group there appeared to be a preference for the substitution of the Thr/Ser at position 4 by a Pro (additional file 3). In the putative EBPs of *Fusobacterium nucleatum nucleatum* ATCC 25586 and *Jannaschia* sp. CCS1, Thr/Ser at position 4 is replaced by Glu. Considering the fact that both organisms lack Sigma-54, the observed replacement may have resulted in a Sigma-70 dependency, like observed for DhaR and TyrR. The fact that Fusobacterial EBP is orthologous to PhhR of *Pseudomonas aeruginosa*, a paralog of TyrR which was shown to be able to replace TyrR as a repressor of the *aroF-tyrA* operon in *E. coli* [78], supports this assumption.

Previous comparative analyses of the Sigma-54 associated EBP[54]s made clear that the Sigma-54 activators connect to a wide spectrum of input signals [15,52]. In fact, this could be one of the main reasons that a common functional theme thus far has remained obscured. Currently, the PFAM database describes around 136 different domain compositions (architectures) for EBP[54]s. Nevertheless, some generalizations can be made. All EBP[54]s possess a central activator domain, which is responsible for the interaction with Sigma-54 and provides the ATPase activity that is required to initiate transcription. In addition, most enhancer-binding proteins have one to several signal binding/recognition domains and a DNA-binding domain, although some EBP[54]s lack either the former or the latter [18,44,79,80]. Basically two main domain organizations can be discerned, which can be further subdivided according to specific domain composition as indicated schematically in Figure 3B.

To specify the functional associations of the Sigma-54 activators, PFAM domain annotations were collected from the Uniprot database and used to make an inventory of the specific EBP[54] categories. We found that around 43% of the activators were of type Ia, thus related to the transduction of extracellular signals/cues via a two-component system histidine kinase (for review on TCSs see [81,82]). Furthermore, around 12% of the activators contained a PAS domain and around 10% a GAF domain (both type Ib). Earlier comparative analyses have shown that these two are the most wide-spread domains among bacterial signaling systems [83,84]. The PAS domain has been linked to a variety of (small) signaling molecules [85] and the same holds for the GAF
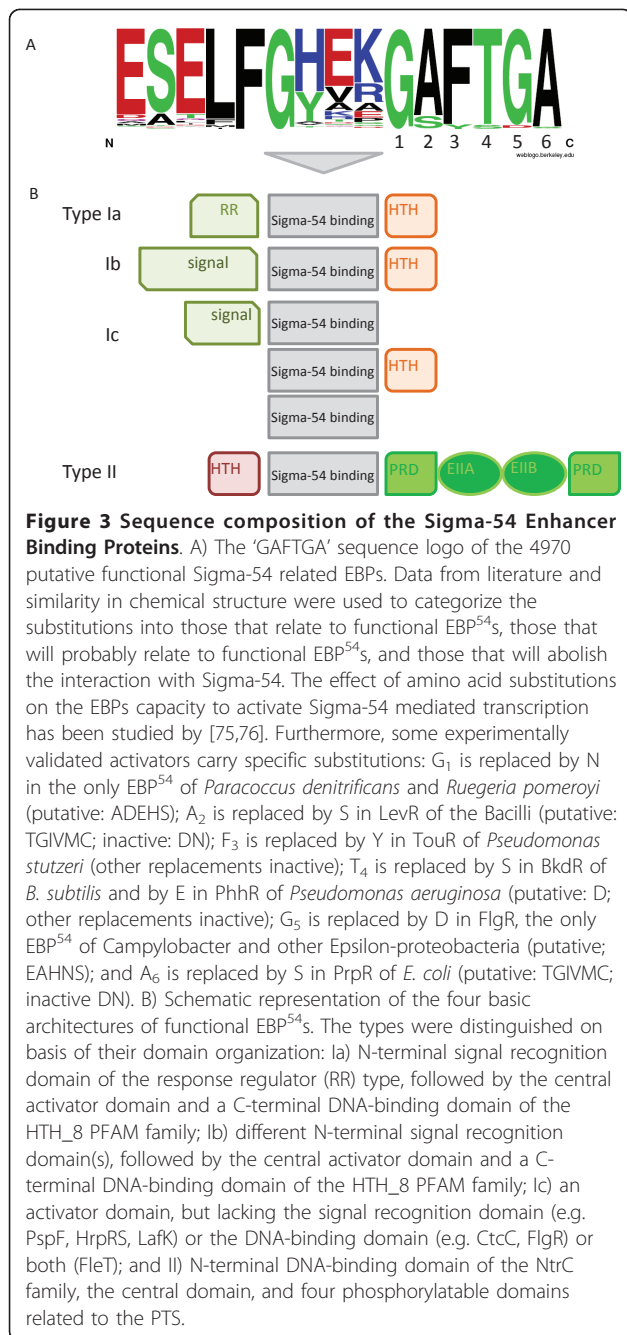
domain (see [38]), although it was originally linked to the recognition of cyclic nucleotides (see [86]). In several cases the PAS and GAF domain occurred in conjunction with another signal-recognition domain, which is suggestive of dual activation and or signal integration. Around 4% of the activators was of type Ic, lacking a DNA-binding domain and/or signal recognition domain. Finally, around 3% of the activators was directly linked to signaling via the PTS. These activators included proteins of type Ib with an HPr signal-domain found in Clostridia (0.3%) [87] and the proteins constituting type II (2.5%), with four different phosphorylatable domains (2xEII and 2xPRD; see [37]). Notably, a substantial number of EBP[54]s lacked a clear PFAM annotation of the putative signal recognition domain. Given the specificity of the Hidden Markov Models used to identify the response regulator domain one may assume that these activators mainly represented types 1b and 1c. Around 8% of all EBP[54] sequences was shorter than 400 amino acids, making them likely of type 1c. Based on the above, type 1b activators represented around 46% of the EBP[54]s identified.

More than half of the activators (~60%) could be annotated in a coarse way, on basis of the similarity to experimentally verified EBP[54]s. As expected, the coverage within the Gamma-proteobacteria and the Firmicutes was higher than for other classes. Nevertheless, the data clearly indicate that the 'membership' and cross-phylum conservation is much higher for certain annotations than for others, as illustrated in Table 1. More explicitly, the annotations within the largest group include connections to the catabolism of short-chain fatty-acids and chemotaxis (AtoC), the synthesis of lipoprotein (Rrp2) and the response to cyclic-di-GMP (e.g. ZraR), a second messenger that regulates cell-surface-associated traits [88,89]. Other highly represented and cross-phylum conserved associations include the control of nitrogenase (VnfA, 11 phyla/classes), nitric oxide reduction (NorR), nitrogen assimilation and the production of EPS and LPS (NtrC), the transport and metabolism of (di-)carboxylates (AcoR, DctD, FhlA, GabR, PrpR), flagellar synthesis (AdnA, FleQ, FleR, FlrA, FlrC) and the degradation and uptake of various kinds of cell wall (poly-)saccharides (CelR- and LevR-like). Minor

Table 1 Main classes of Sigma-54 related activators and the connected functional data.

	annotation	nr	% of total	phyla/ class	members	associated processes
Ia)	AtoC-like	408	8.2	16	AtoC, FrgC, HydG, Rrp2, ZraR	AtoC: catabolism of short chain fatty acids induced by **acetoacetate**; biosynthesis of **polyhydroxybutyrate**; related to motility and expression of **flagella**r genes [124]; FrgC: developmental association in *Myxococcus xanthus* [180]; Rrp2: synthesis of **lipoproteins** in Borrelia [181]; ZraR: responds to **cyclic-di-GMP** [182]
	NtrC-like	302	6.1	5	NtrC, NRI	**nitrogen assimilation** in Gamma-proteobacteria (see [25]); **choline catabolism** in *Pseudomonas aeruginosa* [43]; biosynthesis of **EPS** and **LPS** in *V. vulnificus* [183]and *P. aeruginosa* [184];
	DctD-like	209	4.2	3	DctD	(C4) **dicarboxylate** transport, associated with symbiosis in Rhizobia [137]
	QseF-like	117	2.4	3	QseF, YfhJ	production OM protein in response to host **pheromone** or **sulphate** and **phosphate** stress [185]
	FleR-like	86	1.7	4	FleR, FlrC	**motility** and adhesion to mucin in *Pseudomonas aeruginosa* [186]
Ib)	AcoR-like	216	4.3	6	AcoR	**acetoin** metabolism in *B. subtilis* [187]
	NorR-like	146	2.9	3	NorR	**nitric oxide** detoxification in *E. coli* [188]
	PrpR-like	129	2.6	3	PrpR	control of **propionate** metabolism in *Salmonella* [189]
	FhlA-like	91	1.8	4	FhlA, HyfR	FhlA: controls expression of formate hydrogen lyase and is induced by **formate** [190]
	GabR-like	90	1.8	4	GabR	GabR: control of the **gamma amino-butyric acid** shunt in *Bacillus thuringiensis* [40]
	FleQ-like	81	1.6	1	FleQ, AdnA, FlrA	synthesis of **flagella** in *Legionella pneumophila* [115]
	VnfA	51	1.0	11	VnfA	control of **nitrogenase** expression (see [191])
Ic)*	PspF	139	2.8	3	PspF*	control of **membrane stress** response (see [80])
II	LevR-like	123	2.5	3	CelR, LevR	LevR: controls *lev* operon in *B. subtilis*. Operon includes PTS transport of **polyols** and other **sugar derivatives** [130]; CelR: controls *cel* operon in *Geobacillus stearothermophilus*. Operon includes PTS with strong activity towards plant **cell wall carbohydrates** [131]; EsuR: controls *esu* operon, which is related to **acetyl-sugar** uptake and hydrolysis [132] lmo1721: *cel* operon in *Listeria monocytogenes*; control of **virulence** genes [133]

Column one lists the type of EBP[54] and column two a general group annotation. The annotation was based on BLAST scores against a list of 60 experimentally characterized activators (given in additional file 3) and represents around 60% of the identified activators. Columns 3 and 4 provide the (relative) number of activators that are in the group and column 5 gives the number of phyla/classes that are represented (maximum 17). Column 6 lists the characterized activators that are member the group. The final column provides a function description for the specified group member as retrieved from literature, where the main metabolite or process is in bold type.*The EBPs of type Ic, include PspF [80], HrpR (0.3%) [44,176] and LafK (0.3%) [177] which lack a separate signal recognition domain, Chlamydia CtcC [178] and Helicobacter FlgR [179] which lack a DNA-binding domain. and *R. sphaeroides* FleT [117] which lacks both domains.
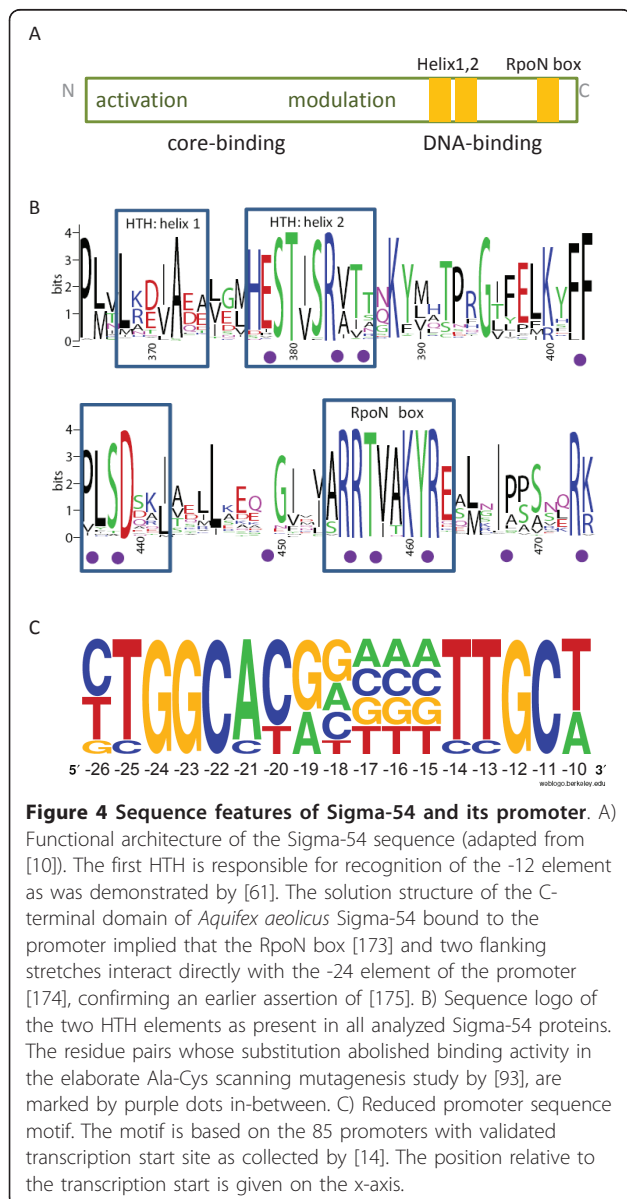
activities not listed include the sensing, transport and metabolism of hydrocarbons (XylR-like [42]), aromatic amino acids (PhhR [78] or CbrB [90]) and the production of the EPS alginate (AlgB [91]).
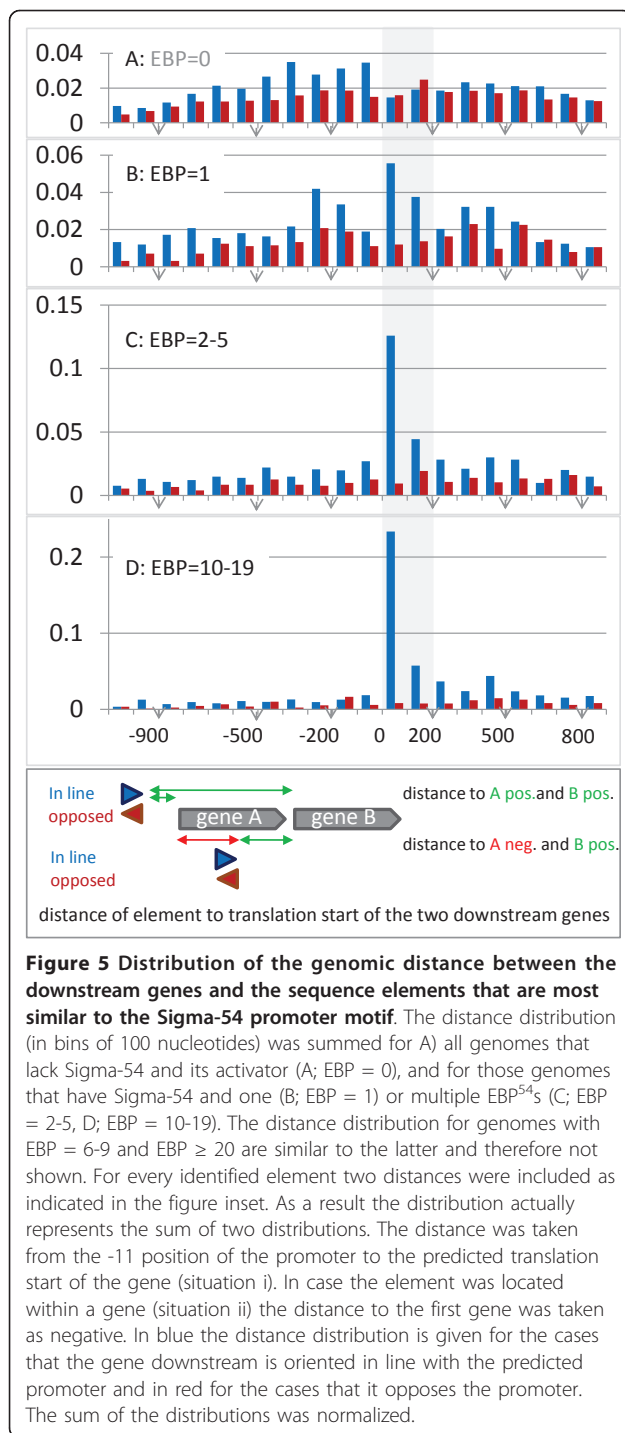
Identification of Sigma-54 promoters

The Sigma-54 amino acid sequence can be divided into three distinct parts that show a high degree of conservation [92] as depicted in Figure 4A. However, the sequences that link the parts can vary considerably depending on the species. The C-terminus contains two separate Helix-Turn-Helix motifs (HTH) that are responsible for recognition of the -24 and -12 promoter elements, respectively. The multiple sequence alignment

Figure 4 Sequence features of Sigma-54 and its promoter. A) Functional architecture of the Sigma-54 sequence (adapted from [10]). The first HTH is responsible for recognition of the -12 element as was demonstrated by [61]. The solution structure of the C-terminal domain of *Aquifex aeolicus* Sigma-54 bound to the promoter implied that the RpoN box [173] and two flanking stretches interact directly with the -24 element of the promoter [174], confirming an earlier assertion of [175]. B) Sequence logo of the two HTH elements as present in all analyzed Sigma-54 proteins. The residue pairs whose substitution abolished binding activity in the elaborate Ala-Cys scanning mutagenesis study by [93], are marked by purple dots in-between. C) Reduced promoter sequence motif. The motif is based on the 85 promoters with validated transcription start site as collected by [14]. The position relative to the transcription start is given on the x-axis.

of all recovered Sigma-54 sequences shows that the amino acid residues of the two HTH motifs are especially well-conserved (Figure 4B). The degree of conservation of the HTH residues corresponds remarkably well to the negative effect on activity that was measured after the substitution of residue pairs [93]. Considering the high degree of conservation within the DNA-binding sequence of Sigma-54, one would expect a similar degree of conservation for the promoter sequence. Indeed, a consensus promoter sequence has been defined earlier by combining experimentally verified promoters from a variety of species [14]. The consensus has been used by many researchers to search for new putative Sigma-54 binding sites (e.g. [94]). Nevertheless, it is very clear from the conservation pattern within the consensus promoter that some positions are not very informative. Therefore, we reduced the size of the promoter motif and smoothed those frequencies that appeared to be rather random, to arrive at a frequency representation of the Sigma-54 promoter as given in Figure 4C.

The reduced Sigma-54 promoter motif was used to identify similar sites in all the studied genomes. To that end, we formulated a straightforward frequency-based scheme to score similarity and implemented the scheme in a similar motif search (SMS) tool (see methods). SMS was tested to predict the well-studied CcpA and Spo0A regulon in *B. subtilis* and the simple scoring appeared as effective as MAST and more effective than HMMs in finding members of the respective regulons (see methods). The results of the similar motif search for the Sigma-54 promoter were evaluated by setting an arbitrary initial score threshold (85% of the maximum obtainable score), and then counting the number of occurrences in every genome. As expected, there was a clear correlation between the number of identified similar sites and genome size. In contrast, there was no clear difference between the number of potential binding-sites in organisms that have Sigma-54 *versus* organisms that do not. However, in case the orientation and the distance of the potential promoters with respect to the predicted translation start sites of the genes located downstream was taken in to account, the difference was obvious (as depicted in Figure 5). For organisms that possess Sigma-54, a large fraction of the most similar binding-sites is located in the region between 0 and 200 nucleotides upstream of a translation start (see Figures 5BCD), whereas in organisms that lack Sigma-54 the distribution of sites shows a slight decrease in this region (Figure 5A). In fact, this finding implies that those potential sites that are appropriately located to function as a promoter are highly likely to function as a genuine Sigma-54 promoter. Therefore, the list of potential sites could be reduced using a simple distance

Figure 5 Distribution of the genomic distance between the downstream genes and the sequence elements that are most similar to the Sigma-54 promoter motif. The distance distribution (in bins of 100 nucleotides) was summed for A) all genomes that lack Sigma-54 and its activator (A; EBP = 0), and for those genomes that have Sigma-54 and one (B; EBP = 1) or multiple EBP^{54}s (C; EBP = 2-5, D; EBP = 10-19). The distance distribution for genomes with EBP = 6-9 and EBP ≥ 20 are similar to the latter and therefore not shown. For every identified element two distances were included as indicated in the figure inset. As a result the distribution actually represents the sum of two distributions. The distance was taken from the -11 position of the promoter to the predicted translation start of the gene (situation i). In case the element was located within a gene (situation ii) the distance to the first gene was taken as negative. In blue the distance distribution is given for the cases that the gene downstream is oriented in line with the predicted promoter and in red for the cases that it opposes the promoter. The sum of the distributions was normalized.

criterion (i.e. -50 - 300 nucleotides upstream of translation start) to yield a list of putative Sigma-54 controlled genes/operons (the results of the promoter identification are available at http://www.cmbi.ru.nl/bamics/supplementary/Franckeetal_2011_Sigma54theme).

The validity of the above identifications was substantiated by a comparison of several predicted species-specific Sigma-54 regulons with those reported in literature. The comparison included *B. subtilis* [95], *E. coli* [25,96], *Lactobacillus plantarum* [23] and *Pseudomonas putida* [53] (see additional file 4). All regulons that were compared showed a very good agreement for the high ranking predicted promoters (i.e. having > 85% of the maximum attainable score). The high-ranked predicted promoters captured more than 95% of the established sites in *Pseudomonas putida*, *B. subtilis* and *Lactobacillus plantarum*. In *E. coli*, 90% of the established Sigma-54 promoters was captured in case a slightly lower threshold was used (i.e. > 80% of maximum). The relatively higher number of less similar 'true' sites in *E. coli* is in line with earlier findings [25,34,96]. Not only did we observe excellent agreement between the predictions and established Sigma-54 promoters, we also identified new likely 'regulon' members. These include for instance: the genes lp_0562 (*nagA*, encoding an N-acetyl-glucosamine-6-phosphate deacetylase) and lp_0586 (*pts10A*, encoding a mannose PTS EIIA) in *Lactobacillus plantarum* WCFS1; the genes PP_0662 (encoding a threonine synthase), PP_4359 (*fliL*, encoding a flagellar basal body-associated protein) and PP_1705 (*nirB*, encoding the large subunit of nitrite reductase) in *Pseudomonas putida*; and the genes b3529 (*yhjK*, encoding a putative diguanylate cyclase, b1786 (*yeaJ*, encoding a diguanylate cyclase involved in the regulation of motility [97]), b2870 (ygeW, encoding an aspartate/ornithine carbamoyltransferase [98] and b4444 (*glmY*, encoding the small RNA that is involved in the activation of expression of glucosamine-6-phosphate synthase [99]) in *E.coli*. Expression of the small RNA encoding gene *glmY* was indeed proven to be Sigma-54 dependent in *E. coli* and other enterobacteria such as *Yersinia pseudotuberculosis* and *Salmonella typhimurium* [99,100].

Conserved function tendencies in genomic context
We collected various kinds of annotations to identify a potential underlying functional theme, for genes: i) that share genome context with the gene encoding Sigma-54 (10 genes upstream and downstream); ii) that occur in transcriptional units containing an EBP^{54}-activator encoding gene (operons and divergons); and iii) that are in transcriptional units preceded by a putative Sigma-54 promoter (see methods for more details). The annotations that were extracted from the reference databases included: COG (av. 73% ± 1%) and GO categories (63% ± 5%), PFAM (83% ± 3%) and Interpro (77% ± 1%) domains, KEGG orthologs (59% ± 5%), EC numbers (9% ± 2%), trivial gene names (22% ± 3%) and detailed function descriptions (67% ± 6%). The average coverage of the various annotations that is given between brackets for the three context collections shows that only part of the recovered genes was connected to annotation

information. The COG categories, PFAM domains and KEGG orthologs together connect more than 80% of the complete set of genes to a function annotation and therefore these annotations should represent the overall genetically associated functions to a reasonable extent. The annotations were lumped in a phylum-specific manner and those annotations that were highly represented within a phylum or class (top 10 or 20, depending on number of genomes) and represented within several phyla/classes (≥ 2), were extracted. The results of the procedure can be found in additional files 5, 6 and 7, respectively, and are summarized in the following.

Table 2 presents an overview of the COG functional categories that were found over-represented in the set of genes that are directly (i.e. genomically) associated with the genes encoding Sigma-54 and its activators and with the identified Sigma-54 promoters. The representation was determined relative to the complete set of proteins in the COG database. There appeared to be six prevalent categories, namely: Energy production and conversion, Cell wall/membrane/envelope biogenesis, cell motility, post-translational modification, signal transduction and intracellular trafficking/secretion. Interestingly, the category 'amino acid transport' and biosynthesis was not over-represented.

We then inspected the recovered annotations more specifically and found that various genes/function descriptions are often genomically associated to Sigma-54 mediated control. Not surprisingly, the main function associations found within the genetic context of the EBP[54]s and the best ranked promoters were very similar to those that have been reported in literature for the various Sigma-54 EBP-activators. The retrieved genes encoded: various histidine kinases, transport systems of metal ions, nitrate or ammonium, carboxylates, amino acids and cell wall (amino-)sugars and many flagellar proteins (*flg, fli, flh* operons). The genetic context of the Sigma-54 encoding gene was remarkably conserved throughout. In the didermal species it contained: *yhbJ*, encoding a regulator of glucosamine-6-phosphate synthase [101], *lptABC*, encoding the system responsible for the transport of LPS from inner- to outer-membrane [102-104], *kdsCD*, encoding genes related to the synthesis of the LPS building block 2-keto-3-deoxyoctanate (kdo) [105], PTS^{Ntr}, a system implicated in the integrative regulation of nitrogen and carbon metabolism [37,38], and *yhbH* (HPF), a gene implicated in the phenomenon of ribosome hibernation that occurs in stationary phase [106]. In the Sigma-54 containing monoderm species (i.e. the Firmicutes), a conserved association was found with *yhbJ*, like in the diderm species, and with the genes encoding the central glycolytic enzymes and their regulator CggR [107].

The highly represented annotations related to metabolic reactions were mapped on a metabolic map,

Table 2 Representation of the COG categories within the genetic context of the genes encoding Sigma-54 and its EBP[54]s and of the promoters

code	Description of category	s54,	EBP,	prom
A	RNA processing and modification	−	−	−
C	**Energy production and conversion**	−	+	+
D	Cell cycle control, cell division, chromosome partitioning	−	−	−
E	Amino acid transport and metabolism	−	−	+
F	Nucleotide transport and metabolism	−	−	−
G	Carbohydrate transport and metabolism	+	−	−
H	Coenzyme transport and metabolism	−	−	−
I	Lipid transport and metabolism	−	+	−
J	Translation, ribosomal structure and biogenesis	+	−	−
K	Transcription	−	−	−
L	Replication, recombination and repair	+	−	−
M	**Cell wall/membrane/envelope biogenesis**	+	+	+
N	**Cell motility**	−	+	+
O	**Posttranslational modification, protein turnover, chaperones**	+	+	+
P	Inorganic ion transport and metabolism	−	−	−
Q	Secondary metabolites biosynthesis, transport and catabolism	−	−	−
T	**Signal transduction mechanisms**	+	+	+
U	**Intracellular trafficking, secretion, and vesicular transport**	+	+	+
V	Defense mechanisms	−	−	−

Over-representation of a particular COG category (with respect to the reference; http://www.ncbi.nlm.nih.gov/COG/old/) within at least half of the analyzed phyla/classes is indicated by +. The categories that were over-represented in more than two genetic contexts are indicated in bold. The functional categories that are not relevant because they are either non-specific (R, S) or mostly specific for eukaryotes (B, W, Y, Z) are not shown.

representing most generic biochemical pathways, using the webtool iPATH [108] (Figure 6). It was immediately clear from the patchy appearance that the retrieved genetically associated functions did not seem to converge on certain predefined metabolic pathways like for instance glycolysis or amino acid biosynthesis. However, when the metabolites that were involved in the mapped reactions were inspected, a surprisingly clear picture emerged. These metabolites included various carboxylates (e.g. glutamate, acetate, propionate, butanoate, formate), charged coenzyme A (which is related to fatty acid synthesis or the TCA cycle), (deoxy-)nucleotides, several nitrogen-containing metabolites including various amino-sugars, and the central glycolytic molecules (e.g. pyruvate, phosphoenolpyruvate, 3-phosphoglycerate and 3-phosphoglycerol). Together, the metabolites constitute the essential precursors for the biosynthesis of EPS (i.e. activated (amino-)sugars [109]), lipoprotein and phospholipids (i.e. 3-phosphoglycerol, fatty acids and amino acids [110]), (lipo-)teichoic acids (i.e. polyols, activated sugars, alanine and ribitol or 3-phosphoglycerol [111]), LPS (i.e. activated amino-sugars, activated C6 and C7 sugars, 2-keto-3-deoxyoctonate (kdo) and fatty acids [112]), and peptidoglycan (activated aminosugars, glutamine, alanine and diaminopimelate or lysine [113]) [114]. Indeed, the essential reactions of the related synthesis routes appeared to be highly represented and conserved. The above findings strongly support a common regulatory role for Sigma-54 in the control of the biosynthesis of the bacterial exterior.

Discussion

We have applied a coherent comparative sequence-based strategy to search for functional themes that are common to Sigma-54 mediated control. The strategy basically consisted of three semi-independent comparative analyses concerning: i) the taxonomic distribution; ii) the Sigma-54 activator content; and iii) the genomic context of Sigma-54, its activator and the characteristic promoter. In principle, the former and the latter analysis can be used to identify the functional theme that is associated to any bacterial regulator.

One of the main challenges we encountered in the analysis of the genetic context of Sigma-54, its EBP[54]-activators and the Sigma-54 promoters, was the fact that most of the function information that is gathered in larger resources in the public domain is given and viewed in terms of a limited set of established biochemical pathways and/or functional classes. As a result, the recovered genetic associations *per se* did not reveal a single over-represented functional category (using COG categories) or a complete pathway (using KEGG orthologs), in line with earlier observations. However, changing the perspective from isolated categories and

pathways to a more integrated (systems) view, and from pathways to compounds, all of a sudden a coherent pattern emerged (see Figure 6). Most of the conserved reactions and transport systems relate directly or indirectly to the biosynthesis of EPS, lipoprotein, (lipo-)teichoic acids, LPS, peptidoglycan and phospholipids, by producing and/or transporting the particular precursors/building blocks. In fact, this common functional theme of being a controller of the synthesis of the bacterial exterior covers very well the diversity within the reported physiological effects of Sigma-54. In the following we will discuss the foundation of this assessment in more detail.

Considerations based on conservation

A clear relation between the activity of Sigma-54 and the synthesis of flagella and or pili has been established in a wide variety of bacterial species (see e.g. [34,35,49,50,115-118] and reviews by [71,119]). Indeed, this association appears very general considering the strong correlation between the presence of Sigma-54 and cellular motility (Figure 2C). Moreover, a comparison of the presence/absence distribution with the transition-based tree of life advocated by [120,121] was in line with this generalization. Those species that were proposed to have developed before the advent of flagella lack Sigma-54, whereas the majority of species that were proposed to have arisen later have the sigma factor.

The classification of the various phyla based on morphological features (see [63]) as presented in Figure 1, revealed another clear functional association. Sigma-54 is mainly present in diderm organisms that synthesize LPS. Although the association is not absolute, a closer inspection of the diderm species that lack Sigma-54 showed that more than 90% of them are (facultative) endosymbionts. The organisms of the phylum Cyanobacteria constitute a notable exception. However, it has been reported that their LPS is distinct from that of the other LPS-containing diderm species [122]. Thus, a straightforward comparison of the morphology and mobility of species that have Sigma-54 versus species that do not, provides a clear link between the sigma factor and the presence of flagella and the biosynthesis of LPS. In fact it was shown recently, that the assembly of the flagella and the LPS in *Campylobacter jejuni* is an associated process and linked through a single enzyme [123].

We have collected a complete set of EBP[54]s for the analyzed genomes and classified them on basis of domain structure and similarity to experimentally verified activators. The annotation of the main groups of EBP[54]s suggested a clear relation to the biosynthesis of the bacterial exterior and the transport and synthesis of the required precursors. The most dominant activator
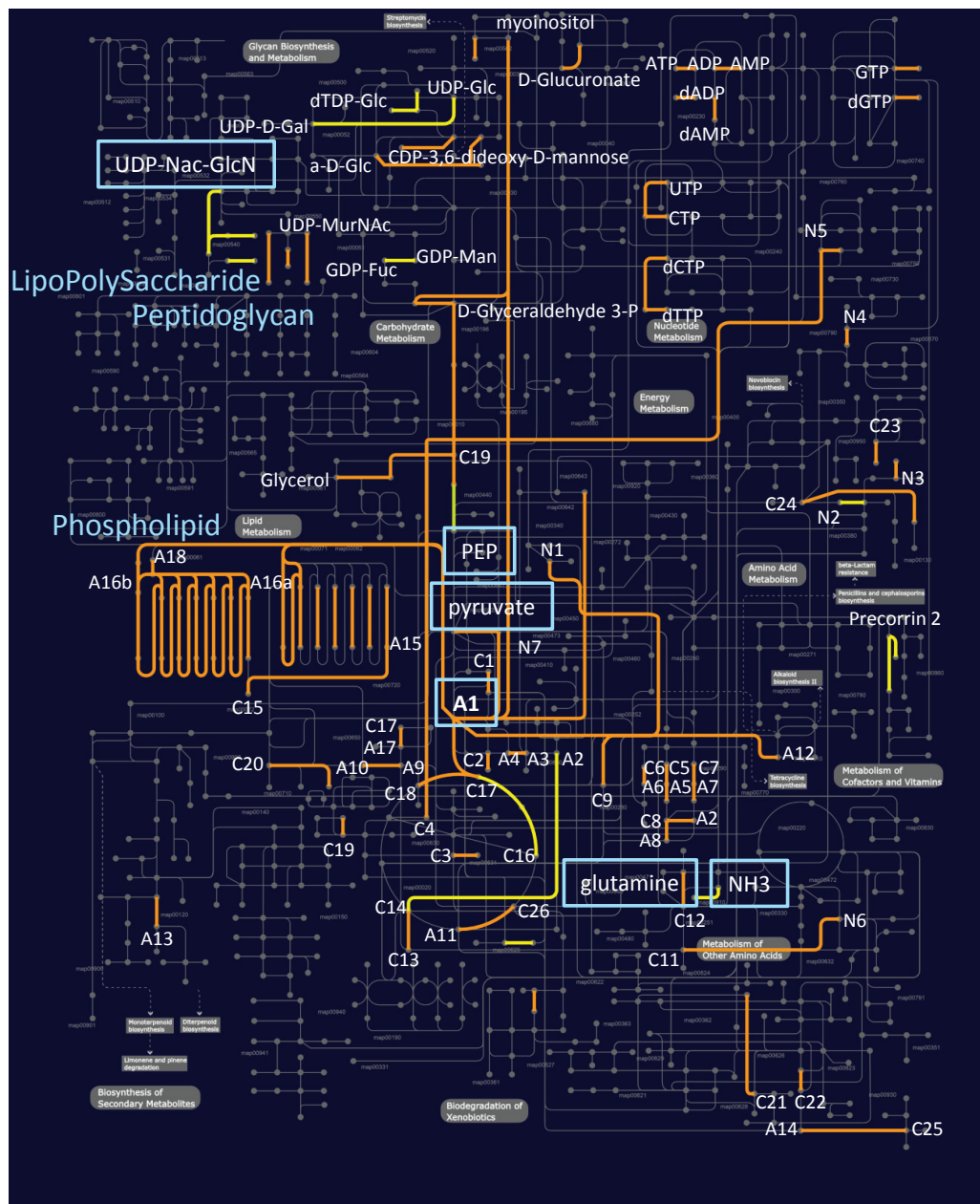
Figure 6 Conserved function tendencies within the gene-associations of Sigma-54, its EBP[54]s and the Sigma-54 promoter. The highly represented and cross-phylum conserved metabolic reactions were mapped using iPATH [108]. The reactions that relate to only Firmicutes are colored green, those that relate to diderm organisms only are colored yellow and those reactions represented in both monoderm and diderm species are given in orange. The routes associated with phospholipid, peptidoglycan and lipopolysaccharide biosynthesis are indicated and the related precursors are given in blue boxes. The metabolites that are associated to the recovered reactions fall in 3 distinct categories. i) CoA-related: A1, acetyl-CoA; A2, propanoyl-CoA; A3, propenoyl-CoA; A4, 3-hydroxypropanoyl-CoA; A5, 2-methylpropanoyl-CoA; A6, 3-methylbutanoyl-CoA; A7, 2-methylbutanoyl-CoA; A8, (R)-2-methyl-3-oxopropanoyl-CoA; A9, 2-butenoyl-CoA; A10, (S)-3-hydroxybutanoyl-CoA; A11, succinyl-CoA; A12, glutaryl-CoA; A13, 3alpha,7alpha-dihydroxy-5beta-cholestanoyl-CoA; A14, 3-oxoadipyl-CoA; A15, hexadecanoyl-CoA; A16a, acetoacetyl-CoA; A16b, acetoacetyl-[acp]; A17, butanoyl-CoA. ii) carboxylates: C1, acetate; C2, 3-oxopropanoate; C3 glycolate; C4, malate; C5, 3-methyl-2-oxobutanoate; C6, 4-methyl-2-oxopentanoate; C7, 3-methyl-2-oxopentanoate; C8, (S)-methylmalonate semialdehyde; C9, L-aspartate; C10, butanoate; C11, 4-aminobutanoate; C12, L-glutamate; C13, succinate semialdehyde; C14, succinate; C15, hexadecanoate; C16, isocitrate; C17, citrate; C18, oxaloacetate; C19, 3-phospho-D-glycerate; C20, acetoacetate; C21, salicylate; C22, 3-oxoadipate; C23, 3,4-dihydroxymandelaldehyde; C24, chorismate; C25, 6-oxohexanoate; C26, 2-oxoglutarate. iii) amino-group containing: N1, histamine; N2, anthranilate; N3, 5-hydroxytryptamine; N4, 2-amino-4-hydroxy-6-(erythro-1,2,3-trihydroxypropyl)dihydropteridinetriphosphate; N5, Nicotinate; N6, 1,4-butanediamine; N7, 2-hydroxyethyl-ThPP.

subgroup was formed by the AtoC-like activators (of type Ia) and was represented in all but one of the analyzed phyla. The main family member AtoC has been associated in the literature with the catabolism of short-chain fatty-acids, the biosynthesis of polyhydroxybutyrate and with the expression of flagellar genes [124].

The common functional theme was also perfectly reflected in the genomic context of the Sigma-54 encoding genes. More specifically, in many Firmicutes the enzymes of the lower branch of glycolysis are genetically associated, thus linking Sigma-54 to the synthesis of 3-phosphoglycerate and phosphoenolpyruvate (PEP). The former compound can easily be converted to 3-phosphoglycerol, whereas PEP is essential to convert UDP-N-acetylglucosamine to UDP-N-acetylmuramic acid during the formation of peptidoglycan [113]. In addition, PEP serves as free energy donor in the transport of (amino-)sugars mediated by the PTS. In most diderms the LPS transport system [102-104] and enzymes involved in synthesis of the building block of the LPS core (2-keto-3-deoxyoctonate, kdo) [105] are genetically associated to the Sigma-54 encoding gene. In fact, it has been proposed that this association represents the ancestral state of all Proteobacteria [125]. The same holds for *yhbJ*, a gene that encodes an important regulator of the glucosamine-6-phosphate synthase and is controlled by Sigma-54 in *E. coli* [101], where glucosamine-6-phosphate is the main amino-sugar precursor in the synthesis of LPS and the bacterial cell wall [112-114]. Interestingly, in the enterobacteria, Sigma-54 also controls the activity of glucosamine-6-phosphate synthase via the expression of the small RNA regulators GlmY and GlmZ [99,100].

The direct links to central nitrogen and central carbon metabolism

A specific 'nitrogen-PTS' (PTS^{Ntr}) and mannose-PTS were also proposed to be part of the ancestral Proteobacterial association [125]. The PTS^{Ntr} is found in Proteobacteria (except for the Epsilon-proteobacteria) [126] and is involved in the integrative regulation of carbon and nitrogen metabolism [37,38,127]. In addition, the PTS^{Ntr} has been connected to the control of biofilm formation in *Vibrio cholerae* [128,129]. Remarkably, the mannose-PTS connection is found in all Firmicutes, either directly like in the Lactobacilli *Lactobacillus plantarum* [23] and *Pediococcus pentosaceus*, or indirectly via the EBP^{54}-activators. The latter LevR/CelR-like activators are found in all Sigma-54 containing *Firmicutes* and in some enterobacteria. Their activity is controlled by extracellular carbohydrate and intracellular PEP levels via the PTS [37,59]. The connected PTS systems have been shown to transport cell-wall related carbohydrates [130-133] and belong to the cellobiose and mannose

sub-families. It was recently shown by [134] that the mannose-family PTS ManXYZ of *E. coli* plays an important role in the secretion of the glucosamine intermediates from the cytoplasm to the periplasm where the *E. coli* cell wall is assembled.

Regulation by Sigma-54 is clearly linked to central nitrogen metabolism. In a noteworthy attempt to uncover an underlying functional theme for Sigma-54 mediated transcriptional control within *E. coli* [25] concluded that nitrogen assimilation was a major theme and that, considering the widespread distribution of the Sigma-54 system together with the constraints determined by maintenance of promoter integrity, the role of Sigma-54 should be limited to a few physiologically related themes in the various organisms. In fact, considering the composition of the bacterial exterior (listed earlier) and the way the different components are covalently linked, the association with nitrogen assimilation is rather logical. The important precursors of the biosynthesis of the molecules of the exterior are (N-acetyl-) glucosamine, Glu, Lys, PEP and various fatty acids (i.e. carboxylates). Most other precursors can be synthesized out of these. For instance, Glu and PEP can be converted to oxaloacetate and Ala [112], and Lys can be interconverted to diaminopimelate (dap) [135]. In the last case, the biosynthesis route starts at aspartate, the transport of which is controlled by the Sigma-54 dependent two-component system DctBD in many species [136,137]. In addition, many of the moieties are finally covalently linked through peptide bonds, which involve an amino- and a carboxyl-group. Therefore, in order to control the direction and scale of the metabolic fluxes related to the biosynthesis of the different exterior structures the bacterium should control the extent of nitrogen assimilation and the production of the central metabolites Glu and PEP. Nevertheless, the way in which the control is exerted may vary between species. For instance, Sigma-54 is controlling the intracellular Glu levels via glutamate dehydrogenase in the monoderm *B. subtilis* [138], whereas it acts on glutamine synthetase in the diderm *E. coli* (see [25]).

Conclusion

We have established a clear connection between Sigma-54 and the make-up of the bacterial exterior. The sigma factor exerts its control directly by regulating the expression of the genes involved in the transport and biosynthesis of the main precursors. In some cases, the control is indirect and mediated via an additional regulator like in the case of PTS^{Ntr} [38] or via another sigma factor like Sigma-S in Borrelia [46] or Sigma-32 in enterobacteria [25]. However, experimental evidence for the latter connection has not yet been reported. In contrast to the situation in Borrelia, in *E. coli*, Sigma-54 and

Sigma-S have been predicted to act in an antagonistic way [139]. Of course, the control exerted by Sigma-54 should not be viewed as an isolated process or acting at the level of transcription alone. Regarding the latter, there seems to be a connection to control at the level of translation activity (e.g. ribosome hibernation in stationary phase [106]). Regarding the former, because Sigma-54 mediated control will affect various fluxes involving central metabolites, other global regulatory factors bear upon its activity, such as the DNA-bending proteins IHF and CRP, the alarmone ppGpp and the RNA-polymerase targeting protein DksA (reviewed in [18]).

Although the association with the exterior constitutes the evolutionary conserved (i.e. pan-bacterial) functional basis of Sigma-54 mediated control, the extent to which the related processes are controlled will very much depend on the species, as is obvious from the variation in the number of Sigma-54 activators. In addition, other processes might have become linked too because they feed on the same metabolic intermediates. Vice versa, such connections might also have been lost and there are several examples of that. These include the activators DhaR, TyrR and VpsR, which are related to the regulation of the uptake and phosphorylation of dihydroxyacetone [28], of the biosynthesis of aromatic amino acids biosynthesis [27], and of the biosynthesis of a particular EPS in Vibrio cholerae [77], respectively. These activators clearly descended from active Sigma-54 dependent EBP[54]s but now carry a small deletion that has made them Sigma-54 independent.

The regulatory connection to the bacterial exterior explains all of the reported physiological variability related to Sigma-54 very well. For instance, the impact on osmotolerance that was observed for *Listeria monocytogenes* upon deletion of Sigma-54 [39] could very well be related to its role in petidoglycan synthesis. In fact, in *E. coli* turgor pressure is controlled through potassium ion transport via the *kdp* system, a system which in turn is controlled via the PTS[Ntr] and thus by Sigma-54 [140]. The reported changes in virulence can also be explained perfectly through changes in the LPS composition. In addition, alterations in the motility and the presence of flagella will directly affect the invasive power, as observed for *Borrelia burgdorferi* [47], and also the swarming behavior on surfaces [141]. Related to that, adaptations of the bacterial exterior mediated by Sigma-54 will clearly have to affect the formation and the properties of bacterial biofilms. And this is precisely what has been observed for many species [142]. Sigma-54 and the associated activators thus represent potentially highly effective targets in the areas of food safety and health as changes in the bacterial exterior induce the establishment and affect the stability of deleterious bacterial populations.

Methods

External Data and Tools

Genome sequence and annotation information was obtained from NCBI [143]. For all species with a sequenced genome that was published before November 2009, the taxonomic attributes and physiological data were derived from the GOLD database [60]. Other function annotation of genes was automatically collected on basis of gi-IDs from the Uniprot database (PFAM, COG, GO and Interpro) [144] and the KEGG database (Kegg ortholog, pathway, linked reactions and compounds) [145]. Sequence similarity searches were performed using BlastP or tBlastN [146], Hidden Markov Models (implemented according to [147]), or using Similar Motif Search (see below). The latest version of ClustalX [148] was used for multiple sequence alignments and for the generation of Neighbor Joining (NJ-) trees (bootstrapped and corrected for multiple substitutions). NJ-trees were visualized and organized using LOFT [149] or Dendroscope [150]. Frequency representations of aligned sequences were created with Weblogo [151]. Wordles [152] were employed to compare the frequency of annotations. The tool iPATH [108] was used to visualize the metabolic context of recovered sets of annotations. All relevant data has been made publicly available at http://www.cmbi.ru.nl/bamics/supplementary/Franckeetal_2011_Sigma54theme.

Similar Motif Scoring (SMS)

The identification of stretches of DNA, RNA or protein sequence with a certain function relies on knowledge of other sequences carrying that particular function and a scoring method to characterize the similarity between the target and the query. In general, sequence comparison algorithms evaluate the statistical relevance of the overlap between a target and a given query. Although such an approach is very powerful for larger sequences, it is less discriminative for smaller sequences, like transcription factor binding sites. Therefore, much effort has been put in the development of advanced scoring methods in the field of DNA-binding site identification [153,154]. Nevertheless, ultimately the current tools provide scores and associated rankings that reflect probability rather than similarity.

Most scoring methods that rely on a known aligned set of input sequences create a position weight matrix (i.e. motif) [56,155,156]. Considering the fact that the number of input sequences is normally limited, a proper sampling of the query sequence space is not provided and thus a probabilistic scoring by default will be skewed. In addition, most methods have to introduce artificial scores in case a specific nucleotide is fully conserved or completely absent at a certain position within the input set [157]. Remarkably, a potential solution to

both problems has implicitly been provided by many researchers who have compared binding-site predictions with experimentally observed changes in transcript levels. One of the most common practices to reconcile prediction with experiment is to minimize the number of differences between the target and the query (or the 'consensus'). In fact, this criterion can be captured in a straightforward scoring using only the position frequency matrix:

Given any number m (≥ 1) of input sequences of size i, the nucleotide frequency $f_{N(j)}$ (where N $\in$ A, C, T, G; and frequency is in terms of fraction) at every position $j = 1$ to i can be used directly to provide all target sequences of size i with a score by adding up the input-based frequencies that relate to the nucleotide composition of the target. Division of the score by the length of the sequence i results in a relative 'similarity' score that can range from 0 to 1.

In case the input sequences are representative for high-affinity sites, the ranking of target sequences according to score should approximately correspond to a ranking based on affinity. In addition, the degeneracy of the input motif can directly be deduced from the scores of the individual input sequences and their deviation from 1.

The simple scoring method was implemented in a Similar Motif Search web-tool within the FG-Web framework (van Hijum unpublished, https://trac.nbic.nl/fgweb/) and the effectiveness of the tool was compared to MAST [158] and Hidden Markov Models [147] using the well-studied CcpA [159] and Spo0A [160] regulons in *B. subtilis* as a test case. These test data-sets were chosen because they relate to global regulators (most difficult to predict regulons) and they were experimentally validated. The test indicated that our SMS method is at least as good to identify putative regulatory elements on basis of known input motifs as the commonly used tool MAST [158] (see additional file 8). In fact, within this 'new' similarity scoring no assumptions have to be made, other than that the given input set represents the sequence one is looking for. Moreover, the SMS method was easily adapted to search specific amino acid motifs in sets of proteins and was successfully used to separate true EBP[54]s from homologs not related to Sigma-54.

Promoter identification
The 85 experimentally verified promoters collected by [14] were used to formulate an initial Sigma-54 promoter sequence motif. The promoter motif was then reduced by smoothing those frequencies that appeared to be rather random, to arrive at a frequency

representation of the Sigma-54 promoter as given in Figure 4C. Although the consensus promoter was valid for species that have a single copy of Sigma-54, it should not necessarily be the case for species that carry multiple copies of the sigma factor. These species include *Bradyrhizobium japonicum* [66], *Rhizobium etli* [67], *Rhodobacter sphaeroides* [69], *Xanthomonas campestris* [68], and a few others. It was reported that the promoters related to the paralogs RpoN1 and RpoN2 in *Rhodobacter sphaeroides* vary only at nucleotide position -11 [70]. Similarly, for *Bradyrhizobium japonicum* two promoter consensus sequences were identified that differed only at position -12 [161]. Considering the fact that the observed natural promoter variability involves single nucleotide substitutions, we assume that our promoter predictions are also relevant for the species with multiple Sigma-54 copies. Indeed, for *Rhodobacter sphaeroides* and *Bradyrhizobium japonicum* the high-ranking putative promoters included representatives of both reported promoters (not shown).

Identification of Sigma-54 and the related EBPs
The sequences of the experimentally characterized Sigma-54 of *E. coli* and *B. subtilis* and 34 experimentally characterized EBP[54]s (see additional file 3 for sequences and references) were used to identify homologs in all sequenced prokaryotic species (data from NCBI non-redundant protein database as of 1st November 2009). In the case of Sigma-54 no selection was needed. In the case of the EBPs a selection of true positives was performed in three steps. First, all hits were filtered on basis of a maximum product of all 34 e-values (< 1e[-5]) and then on basis of a maximum sum of all 34 e-values (< 1). To remove remaining false positives the presence of the characteristic 'GAFTGA' amino acid sequence that is essential for the interaction between the activator and the sigma factor [74] was analyzed (as discussed in the main text). A comparison between the number of genes encoding a true Sigma-54 activator as identified by us and the number of genes reported in the Uniprot database implies that the Uniprot database contains a relatively large number of false-positive identifications. For species lacking Sigma-54, the Uniprot data suggests that 20% has an EBP[54]. Moreover, for species that have Sigma-54 the total number of EBP[54]s listed within the Uniprot database is only correct in 42% of the cases (see additional file 1). Many of the false identifications relate to the lack of filtering on basis of the integrity of the 'GAFTGA' element. Unfortunately, these false identifications extend into the literature. For instance *E. coli* DhaR [25], *V. cholerae* VpsR [77] and *B. subtilis* YplP [24], have been mistakenly ranked as Sigma-54 activators, whereas they clearly lack the 'GAFTGA' element (additional file 3).

Implementation of Context analysis

The gi-IDs of the putative Sigma-54 and EBP[54] proteins and the genomic position of the Sigma-54 related promoters were used to collect annotation information related to neighboring genes from the reference databases (i.e. NCBI, Uniprot, KEGG and GOLD). In case of the promoters, only those that were correctly oriented and spaced (-50 to 300 nucleotides from translation start) and were within the top 15 ranked identified promoters, were used. The retrieval of annotation information associated to the genomic context was encoded in Python and the code was then converted into a local 'context connector' web-tool within the FG-Web framework (van Hijum unpublished, https://trac.nbic.nl/fgweb/). The tool allowed the definition of constraints related to: i) the number of genes neighboring the query gene that should be taken into account; or ii) the organization of those genes, i.e. in 'operons' or 'divergons' (defined using a distance criterion of maximally 300 nucleotides between genes).

Additional material

Additional file 1: Presence-Absence analysis for genes encoding Sigma-54 and the related Enhancers. The file contains: sheet 1 with the results of the presence analysis of Sigma-54 and the related Enhancers in all bacterial genomes published before November 2009 listed in the GOLD database and with NCBI NC-code; and sheet 2, which provides a summary of the presence and absence in relation to genome size.

Additional file 2: All collected Sigma-54 related EBPs with annotation. The file contains: sheet 1, which gives the complete list of putative Sigma-54 related EBPs for the studied genomes (given in additional file 1) in order of product of BLAST e-values related to experimentally verified EBPs (given in additional file 3); sheet 2, which provides the putative annotations; and sheet 3, which gives the annotation summary.

Additional file 3: List of experimentally validated EBPs and a multiple sequence alignment of the 'GAFTGA' region. The file contains: sheet 1, which provides the list of EBPs that was used to search EBP homologs with BLASTP; sheet 2, which gives the list of EBPs that was used to annotate EBP homologs with reference to relevant experimental data in second column; sheet 3, which shows a snapshot of the multiple sequence alignment of the EBPs from sheet two in the GAFTGA region; sheet 4, which gives the list of EBPs with a defective GAFTGA sequence; and sheet 5, providing a summary of the related numbers.

Additional file 4: Promoter predictions for model organisms. The file contains: the promoter predictions for *Bacillus subtilis* (sheet 1), *Escherichia coli* (sheet 2), *Lactobacillus plantarum* (sheet 3) and *Pseudomonas putida* (sheet 4); and a summary of the predictions (sheet 5).

Additional file 5: Representation of the gene annotations in the context of the Sigma-54 encoding genes. The file contains: sheet 1, giving the annotations present in the context of the Sigma-54 encoding genes in all species (10 genes upstream and 10 genes downstream); and a summary sheet 2.

Additional file 6: Representation of the gene annotations in the context of the EBP encoding genes. The file contains: sheet 1, giving the annotations present in the context of the operons and divergons that contain a gene encoding an EBP (max 10 genes upstream and 10 genes downstream and max 300 nucleotides intergenic distance); and a summary sheet 2.

Additional file 7: Representation of the gene annotations in the context of the Sigma-54 promoters. The file contains: sheet 1, giving the annotations present in the genes and operons that follow a correctly spaced and oriented Sigma-54 promoter (member of top 15 ranked hits) (1 gene upstream and max 10 genes downstream); and a summary sheet 2.

Additional file 8: Validation of the Similar Motif Search Procedure. The file contains: sheets 1, 2 and 3, with the outcome of a comparison between SMS and MAST in the identification of CcpA and Spo0A regulon members in *Bacillus subtilis*; and sheet 4 with a summary of the comparison.

List of abbreviations

COG: Cluster of Orthologous Genes; CW: Cell Wall; dap: diaminopimelate; EBP: Enhancer Binding Protein; EC: Enzyme Code; ECF sigma factor: ExtraCytoplasmic Function sigma factor; EPS: ExoPolySaccharide; GO: Gene Ontology; HMM: Hidden Markov Model; HTH: Helix Turn Helix; ID: Identifier; kdo: 2-keto-3-deoxyoctonate; LPS: LipoPolySaccharide; NJ: Neighbor Joining; OM: Outer Membrane; PEP: PhosphoEnolPyruvate; PTS: PhosphoTransferase System; SMS: Similar Motif Search; TCA: TriCarboxylic Acid; TCS: Two Component System.

Author details

[1]TI Food and Nutrition, P.O.Box 557, 6700AN Wageningen, The Netherlands. [2]Kluyver Centre for Genomics of Industrial Fermentation, Julianalaan 67 2628 BC Delft, The Netherlands. [3]Wageningen University and Research Center, Laboratory of Microbiology, Dreijenplein 10, 6703 HB Wageningen, the Netherlands. [4]Netherlands Bioinformatics Centre, Geert Grooteplein 28 6525 GA Nijmegen, The Netherlands. [5]Wageningen University and Research Center, Food and Biobased Research, PO Box 17, 6700 AA The Netherlands. [6]Center for Molecular and Biomolecular Informatics (260), NCMLS, Radboud University Nijmegen Medical Center, P.O.Box 9101, 6500HB Nijmegen, The Netherlands.

Authors' contributions

CF conceived, designed and coordinated the study, carried out motif and functional analyses and drafted and revised the manuscript. TGK conceived and designed the motif search algorithm and carried out motif analysis and helped revising the manuscript; YH implemented the motif search algorithm and validated the algorithm; LO conceived and designed the tool to collect context information and helped revising the manuscript; VS implemented the tool to collect context information and validated the tool; RM and RJS coordinated the study and helped drafting and revising the manuscript. All authors read and approved the final manuscript.

References

1. Gruber TM, Gross CA: **Multiple sigma subunits and the partitioning of bacterial transcription space.** *Annu Rev Microbiol* 2003, **57**:441-466.
2. Kolter R, Siegele DA, Tormo A: **The stationary phase of the bacterial life cycle.** *Annu Rev Microbiol* 1993, **47**:855-874.
3. Hengge R: **Proteolysis of sigmaS (RpoS) and the general stress response in *Escherichia coli*.** *Res Microbiol* 2009, **160**(9):667-676.
4. Zhao K, Liu M, Burgess RR: **The global transcriptional response of *Escherichia coli* to induced sigma 32 protein involves sigma 32 regulon**

activation followed by inactivation and degradation of sigma 32 in vivo. *J Biol Chem* 2005, **280**(18):17758-17768.

5. Wagner MA, Zahrl D, Rieser G, Koraimann G: Growth phase- and cell division-dependent activation and inactivation of the {sigma}32 regulon in *Escherichia coli*. *J Bacteriol* 2009, **191**(5):1695-1702.

6. Kroos L: The *Bacillus* and *Myxococcus* developmental networks and their transcriptional regulators. *Annu Rev Genet* 2007, **41**:13-39.

7. Hecker M, Pané-Farré J, Völker U: SigB-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria. *Annu Rev Microbiol* 2007, **61**:215-236.

8. de Been M, Francke C, Siezen RJ, Abee T: Novel {sigma}B regulation modules of Gram-positive bacteria involve the use of complex hybrid histidine kinases. *Microbiology* 2011, **157**:3-12.

9. Staroń A, Sofia HJ, Dietrich S, Ulrich LE, Liesegang H, Mascher T: The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) sigma factor protein family. *Mol Microbiol* 2009, **74**(3):557-581.

10. Buck M, Gallegos MT, Studholme DJ, Guo Y, Gralla JD: The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. *J Bacteriol* 2000, **182**(15):4129-4136.

11. Kill K, Binnewies TT, Sicheritz-Ponten T, Willenbrock H, Hallin PF, Wassenaar TM, Ussery DW: Genome update: sigma factors in 240 bacterial genomes. *Microbiology* 2005, **151**(Pt 10):3147-3150.

12. Morett E, Buck M: In vivo studies on the interaction of RNA polymerase-sigma 54 with the *Klebsiella pneumoniae* and *Rhizobium meliloti nifH* promoters. The role of NifA in the formation of an open promoter complex. *J Mol Biol* 1989, **210**(1):65-77.

13. Buck M, Cannon W: Specific binding of the transcription factor sigma-54 to promoter DNA. *Nature* 1992, **358**(6385):422-424.

14. Barrios H, Valderrama B, Morett E: Compilation and analysis of sigma(54)-dependent promoter sequences. *Nucleic Acids Res* 1999, **27**(22):4305-4313.

15. Morett E, Segovia L: The sigma 54 bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. *J Bacteriol* 1993, **175**(19):6067-6074.

16. Studholme DJ, Buck M: The biology of enhancer-dependent transcriptional regulation in bacteria: insights from genome sequences. *FEMS Microbiol Lett* 2000, **186**(1):1-9.

17. Wigneshweraraj S, Bose D, Burrows PC, Joly N, Schumacher J, Rappas M, Pape T, Zhang X, Stockley P, Severinov K, *et al*: Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor. *Mol Microbiol* 2008, **68**(3):538-546.

18. Shingler V: Signal sensory systems that impact sigma 54-dependent transcription. *FEMS Microbiol Rev* 2011, **35**(3):425-440.

19. Weiss DS, Batut J, Klose KE, Keener J, Kustu S: The phosphorylated form of the enhancer-binding protein NTRC has an ATPase activity that is essential for activation of transcription. *Cell* 1991, **67**(1):155-167.

20. Batchelor JD, Doucleff M, Lee CJ, Matsubara K, De Carlo S, Heideker J, Lamers MH, Pelton JG, Wemmer DE: Structure and regulatory mechanism of *Aquifex aeolicus* NtrC4: variability and evolution in bacterial transcriptional regulation. *J Mol Biol* 2008, **384**(5):1058-1075.

21. Batchelor JD, Sterling HJ, Hong E, Williams ER, Wemmer DE: Receiver domains control the active-state stoichiometry of *Aquifex aeolicus* sigma54 activator NtrC4, as revealed by electrospray ionization mass spectrometry. *J Mol Biol* 2009, **393**(3):634-643.

22. Mathews SA, Timms P: Identification and mapping of sigma-54 promoters in *Chlamydia trachomatis*. *J Bacteriol* 2000, **182**(21):6239-6242.

23. Stevens MJ, Molenaar D, de Jong A, De Vos WM, Kleerebezem M: sigma54-Mediated control of the mannose phosphotransferase sytem in *Lactobacillus plantarum* impacts on carbohydrate metabolism. *Microbiology* 2010, **156**(Pt 3):695-707.

24. Wiegeshoff F, Beckering CL, Debarbouille M, Marahiel MA: Sigma L is important for cold shock adaptation of *Bacillus subtilis*. *J Bacteriol* 2006, **188**(8):3130-3133.

25. Reitzer L, Schneider BL: Metabolic context and possible physiological themes of sigma(54)-dependent genes in *Escherichia coli*. *Microbiol Mol Biol Rev* 2001, **65**(3):422-444.

26. Jakobsen JS, Jelsbak L, Welch RD, Cummings C, Goldman B, Stark E, Slater S, Kaiser D: Sigma54 enhancer binding proteins and *Myxococcus xanthus* fruiting body development. *J Bacteriol* 2004, **186**(13):4361-4368.

27. Pittard J, Camakaris H, Yang J: The TyrR regulon. *Mol Microbiol* 2005, **55**(1):16-26.

28. Bächler C, Schneider P, Bähler P, Lustig A, Erni B: *Escherichia coli* dihydroxyacetone kinase controls gene expression by binding to transcription factor DhaR. *EMBO J* 2005, **24**(2):283-293.

29. Bowman WC, Kranz RG: A bacterial ATP-dependent, enhancer binding protein that activates the housekeeping RNA polymerase. *Genes Dev* 1998, **12**(12):1884-1893.

30. Davies KM, Lowe ED, Venien-Bryan C, Johnson LN: The HupR receiver domain crystal structure in its nonphospho and inhibitory phospho states. *J Mol Biol* 2009, **385**(1):51-64.

31. Garcia E, Bancroft S, Rhee SG, Kustu S: The product of a newly identified gene, *glnF*, is required for synthesis of glutamine synthetase in *Salmonella*. *Proc Natl Acad Sci USA* 1977, **74**(4):1662-1666.

32. Hirschman J, Wong PK, Sei K, Keener J, Kustu S: Products of nitrogen regulatory genes *ntrA* and *ntrC* of enteric bacteria activate *glnA* transcription in vitro: evidence that the *ntrA* product is a sigma factor. *Proc Natl Acad Sci USA* 1985, **82**(22):7525-7529.

33. Kustu S, Santero E, Keener J, Popham D, Weiss D: Expression of sigma 54 (ntrA)-dependent genes is probably united by a common mechanism. *Microbiol Rev* 1989, **53**(3):367-376.

34. Zhao K, Liu M, Burgess RR: Promoter and regulon analysis of nitrogen assimilation factor, sigma54, reveal alternative strategy for *E. coli* MG1655 flagellar biosynthesis. *Nucleic Acids Res* 2010, **38**(4):1273-1283.

35. Leang C, Krushkal J, Ueki T, Puljic M, Sun J, Juárez K, Núñez C, Reguera G, DiDonato R, Postier B, *et al*: Genome-wide analysis of the RpoN regulon in *Geobacter sulfurreducens*. *BMC Genomics* 2009, **10**:331.

36. Arous S, Buchrieser C, Folio P, Glaser P, Namane A, Hébraud M, Héchard Y: Global analysis of gene expression in an *rpoN* mutant of *Listeria monocytogenes*. *Microbiology* 2004, **150**(Pt 5):1581-1590.

37. Deutscher J, Francke C, Postma PW: How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev* 2006, **70**(4):939-1031.

38. Pflüger-Grau K, Görke B: Regulatory roles of the bacterial nitrogen-related phosphotransferase system. *Trends Microbiol* 2010, **18**(5):205-214.

39. Okada Y, Makino S, Okada N, Asakura H, Yamamoto S, Igimi S: Identification and analysis of the osmotolerance associated genes in *Listeria monocytogenes*. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* 2008, **25**(9):1089-1094.

40. Zhu L, Peng Q, Song F, Jiang Y, Sun C, Zhang J, Huang D: Structure and regulation of the *gab* gene cluster, involved in the gamma-aminobutyric acid shunt, are controlled by a sigma54 factor in *Bacillus thuringiensis*. *J Bacteriol* 2010, **192**(1):346-355.

41. Mouz S, Coursange E, Toussaint A: *Ralstonia metallidurans* CH34 RpoN sigma factor and the control of nitrogen metabolism and biphenyl utilization. *Microbiology* 2001, **147**(Pt 7):1947-1954.

42. Cases I, de Lorenzo V: Promoters in the environment: transcriptional regulation in its natural context. *Nat Rev Microbiol* 2005, **3**(2):105-118.

43. Massimelli MJ, Sanchez DG, Buchieri MV, Olvera L, Beassoni PR, Schweizer HP, Morett E, Lisa AT: Choline catabolism, sigma(54) factor and NtrC are required for the full expression of the *Pseudomonas aeruginosa* phosphorylcholine phosphatase gene. *Microbiol Res* 2011, **166**(5):380-390.

44. Jovanovic M, James EH, Burrows PC, Rego FG, Buck M, Schumacher J: Regulation of the co-evolved HrpR and HrpS AAA+ proteins required for *Pseudomonas syringae* pathogenicity. *Nat Commun* 2011, **2**:177.

45. Bernard CS, Brunet YR, Gavioli M, Lloubes R, Cascales E: Regulation of type VI secretion gene clusters by sigma54 and cognate enhancer binding proteins. *J Bacteriol* 2011, **193**(9):2158-2167.

46. Smith AH, Blevins JS, Bachlani GN, Yang XF, Norgard MV: Evidence that RpoS (sigmaS) in *Borrelia burgdorferi* is controlled directly by RpoN (sigma54/sigmaN). *J Bacteriol* 2007, **189**(5):2139-2144.

47. Fisher MA, Grimm D, Henion AK, Elias AF, Stewart PE, Rosa PA, Gherardini FC: *Borrelia burgdorferi* sigma54 is required for mammalian infection and vector transmission but not for tick colonization. *Proc Natl Acad Sci USA* 2005, **102**(14):5162-5167.

48. Riordan JT, Tietjen JA, Walsh CW, Gustafson JE, Whittam TS: Inactivation of alternative sigma factor 54 (RpoN) leads to increased acid resistance, and alters locus of enterocyte effacement (LEE) expression in *Escherichia coli* O157: H7. *Microbiology* 2010, **156**(Pt 3):719-730.

49. Saldías MS, Lamothe J, Wu R, Valvano MA: *Burkholderia cenocepacia* requires the RpoN sigma factor for biofilm formation and intracellular trafficking within macrophages. *Infect Immun* 2008, **76**(3):1059-1067.

50. Wolfe AJ, Millikan DS, Campbell JM, Visick KL: **Vibrio fischeri sigma54 controls motility, biofilm formation, luminescence, and colonization.** *Appl Environ Microbiol* 2004, **70(4)**:2520-2524.

51. Visick KL: **An intricate network of regulators controls biofilm formation and colonization by Vibrio fischeri.** *Mol Microbiol* 2009, **74(4)**:782-789.

52. Studholme DJ, Dixon R: **Domain architectures of sigma54-dependent transcriptional activators.** *J Bacteriol* 2003, **185(6)**:1757-1767.

53. Cases I, Ussery DW, de Lorenzo V: **The sigma54 regulon (sigmulon) of Pseudomonas putida.** *Environ Microbiol* 2003, **5(12)**:1281-1293.

54. Huynen M, Snel B, Lathe W, Bork P: **Predicting protein function by genomic context: quantitative evaluation and qualitative inferences.** *Genome Res* 2000, **10(8)**:1204-1210.

55. Huynen MA, Gabaldón T, Snel B: **Variation and evolution of biomolecular systems: searching for functional relevance.** *FEBS Lett* 2005, **579(8)**:1839-1845.

56. Rodionov DA: **Comparative genomic reconstruction of transcriptional regulatory networks in bacteria.** *Chem Rev* 2007, **107(8)**:3467-3497.

57. Andersson U, Molenaar D, Rådström P, de Vos WM: **Unity in organisation and regulation of catabolic operons in Lactobacillus plantarum, Lactococcus lactis and Listeria monocytogenes.** *Syst Appl Microbiol* 2005, **28(3)**:187-195.

58. Francke C, Kerkhoven R, Wels M, Siezen RJ: **A generic approach to identify Transcription Factor-specific operator motifs; Inferences for LacI-family mediated regulation in Lactobacillus plantarum WCFS1.** *BMC Genomics* 2008, **9**:145.

59. Lengeler JW, Jahreis K: **Bacterial PEP-dependent carbohydrate: phosphotransferase systems couple sensing and global control mechanisms.** *Contrib Microbiol* 2009, **16**:65-87.

60. Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC: **The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata.** *Nucleic Acids Res* 2010, , **38 Database**: D346-354.

61. Merrick M, Chambers S: **The helix-turn-helix motif of sigma 54 is involved in recognition of the -13 promoter region.** *J Bacteriol* 1992, **174(22)**:7221-7226.

62. Laskos L, Dillard JP, Seifert HS, Fyfe JA, Davies JK: **The pathogenic neisseriae contain an inactive rpoN gene and do not utilize the pilE sigma54 promoter.** *Gene* 1998, **208(1)**:95-102.

63. Gupta RS: **What are archaebacteria: life's third domain or monoderm prokaryotes related to gram-positive bacteria? A new proposal for the classification of prokaryotic organisms.** *Mol Microbiol* 1998, **29(3)**:695-707.

64. Sutcliffe IC: **A phylum level perspective on bacterial cell envelope architecture.** *Trends Microbiol* 2010, **18(10)**:464-470.

65. Geissinger O, Herlemann DP, Mörschel E, Maier UG, Brune A: **The ultramicrobacterium "Elusimicrobium minutum" gen. nov., sp. nov., the first cultivated representative of the termite group 1 phylum.** *Appl Environ Microbiol* 2009, **75(9)**:2831-2840.

66. Kullik I, Fritsche S, Knobel H, Sanjuan J, Hennecke H, Fischer HM: **Bradyrhizobium japonicum has two differentially regulated, functional homologs of the sigma 54 gene (rpoN).** *J Bacteriol* 1991, **173(3)**:1125-1138.

67. Michiels J, Moris M, Dombrecht B, Verreth C, Vanderleyden J: **Differential regulation of Rhizobium etli rpoN2 gene expression during symbiosis and free-living growth.** *J Bacteriol* 1998, **180(14)**:3620-3628.

68. Yang TC, Leu YW, Chang-Chien HC, Hu RM: **Flagellar biogenesis of Xanthomonas campestris requires the alternative sigma factors RpoN2 and FliA and is temporally regulated by FlhA, FlhB, and FlgM.** *J Bacteriol* 2009, **191(7)**:2266-2275.

69. Poggio S, Osorio A, Dreyfus G, Camarena L: **The four different sigma(54) factors of Rhodobacter sphaeroides are not functionally interchangeable.** *Mol Microbiol* 2002, **46(1)**:75-85.

70. Poggio S, Osorio A, Dreyfus G, Camarena L: **Transcriptional specificity of RpoN1 and RpoN2 involves differential recognition of the promoter sequences and specific interaction with the cognate activator proteins.** *J Biol Chem* 2006, **281(37)**:27205-27215.

71. Smith TG, Hoover TR: **Deciphering bacterial flagellar gene regulatory networks in the genomic era.** *Adv Appl Microbiol* 2009, **67**:257-295.

72. Neuwald AF, Aravind L, Spouge JL, Koonin EV: **AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes.** *Genome Res* 1999, **9(1)**:27-43.

73. Erzberger JP, Berger JM: **Evolutionary relationships and structural mechanisms of AAA+ proteins.** *Annu Rev Biophys Biomol Struct* 2006, **35**:93-114.

74. Rappas M, Schumacher J, Beuron F, Niwa H, Bordes P, Wigneshweraraj S, Keetch CA, Robinson CV, Buck M, Zhang X: **Structural insights into the activity of enhancer-binding proteins.** *Science* 2005, **307(5717)**:1972-1975.

75. González V, Olvera L, Soberón X, Morett E: **In vivo studies on the positive control function of NifA: a conserved hydrophobic amino acid patch at the central domain involved in transcriptional activation.** *Mol Microbiol* 1998, **28(1)**:55-67.

76. Zhang N, Joly N, Burrows PC, Jovanovic M, Wigneshweraraj SR, Buck M: **The role of the conserved phenylalanine in the sigma54-interacting GAFTGA motif of bacterial enhancer binding proteins.** *Nucleic Acids Res* 2009, **37(18)**:5981-5992.

77. Mueller RS, Beyhan S, Saini SG, Yildiz FH, Bartlett DH: **Indole acts as an extracellular cue regulating gene expression in Vibrio cholerae.** *J Bacteriol* 2009, **191(11)**:3504-3516.

78. Song J, Jensen RA: **PhhR, a divergently transcribed activator of the phenylalanine hydroxylase gene cluster of Pseudomonas aeruginosa.** *Mol Microbiol* 1996, **22(3)**:497-507.

79. Beck LL, Smith TG, Hoover TR: **Look, no hands! Unconventional transcriptional activators in bacteria.** *Trends Microbiol* 2007, **15(12)**:530-537.

80. Joly N, Engl C, Jovanovic G, Huvet M, Toni T, Sheng X, Stumpf MP, Buck M: **Managing membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to physiology.** *FEMS Microbiol Rev* 2010, **34(5)**:797-827.

81. Hoch JA: **Two-component and phosphorelay signal transduction.** *Curr Opin Microbiol* 2000, **3(2)**:165-170.

82. Cheung J, Hendrickson WA: **Sensor domains of two-component regulatory systems.** *Curr Opin Microbiol* 2010, **13(2)**:116-123.

83. Anantharaman V, Koonin EV, Aravind L: **Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains.** *J Mol Biol* 2001, **307(5)**:1271-1292.

84. Galperin MY: **Bacterial signal transduction network in a genomic perspective.** *Environ Microbiol* 2004, **6(6)**:552-567.

85. Chang C, Tesar C, Gu M, Babnigg G, Joachimiak A, Pokkuluri PR, Szurmant H, Schiffer M: **Extracytoplasmic PAS-like domains are common in signal transduction proteins.** *J Bacteriol* 2010, **192(4)**:1156-1159.

86. Martinez SE, Beavo JA, Hol WG: **GAF domains: two-billion-year-old molecular switches that bind cyclic nucleotides.** *Mol Interv* 2002, **2(5)**:317-323.

87. Reizer J, Schneider B, Reizer A, Saier MH Jr: **A hybrid response regulator possessing a PEP-dependent phosphorylation domain.** *Microbiology* 1999, **145(Pt 5)**:987-989.

88. Pesavento C, Hengge R: **Bacterial nucleotide-based second messengers.** *Curr Opin Microbiol* 2009, **12(2)**:170-176.

89. Schirmer T, Jenal U: **Structural and mechanistic determinants of c-di-GMP signalling.** *Nat Rev Microbiol* 2009, **7(10)**:724-735.

90. Amador CI, Canosa I, Govantes F, Santero E: **Lack of CbrB in Pseudomonas putida affects not only amino acids metabolism but also different stress responses and biofilm development.** *Environ Microbiol* 2010, **12(6)**:1748-1761.

91. Goldberg JB, Dahnke T: **Pseudomonas aeruginosa AlgB, which modulates the expression of alginate, is a member of the NtrC subclass of prokaryotic regulators.** *Mol Microbiol* 1992, **6(1)**:59-66.

92. Merrick MJ: **In a class of its own–the RNA polymerase sigma factor sigma 54 (sigma N).** *Mol Microbiol* 1993, **10(5)**:903-909.

93. Xiao Y, Wigneshweraraj SR, Weinzierl R, Wang YP, Buck M: **Construction and functional analyses of a comprehensive sigma54 site-directed mutant library using alanine-cysteine mutagenesis.** *Nucleic Acids Res* 2009, **37(13)**:4482-4497.

94. Studholme DJ, Buck M, Nixon T: **Identification of potential sigma(N)-dependent promoters in bacterial genomes.** *Microbiology* 2000, **146(Pt 12)**:3021-3023.

95. Sierro N, Makita Y, de Hoon M, Nakai K: **DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information.** *Nucleic Acids Res* 2008, , **36 Database**: D93-96.

96. Gama-Castro S, Jiménez-Jacinto V, Peralta-Gil M, Santos-Zavaleta A, Peñaloza-Spinola MI, Contreras-Moreira B, Segura-Salazar J, Muñiz-Rascado L, Martínez-Flores I, Salgado H, et al: **RegulonDB (version 6.0): gene regulation model of Escherichia coli K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation.** *Nucleic Acids Res* 2008, , **36 Database**: D120-124.

97. Pesavento C, Becker G, Sommerfeldt N, Possling A, Tschowri N, Mehlis A, Hengge R: **Inverse regulatory coordination of motility and curli-mediated adhesion in** *Escherichia coli*. *Genes Dev* 2008, **22**(17):2434-2446.

98. Naumoff DG, Xu Y, Glansdorff N, Labedan B: **Retrieving sequences of enzymes experimentally characterized but erroneously annotated: the case of the putrescine carbamoyltransferase.** *BMC Genomics* 2004, **5**(1):52.

99. Göpel Y, Lüttmann D, Heroven AK, Reichenbach B, Dersch P, Görke B: **Common and divergent features in transcriptional control of the homologous small RNAs GlmY and GlmZ in** *Enterobacteriaceae*. *Nucleic Acids Res* 2011, **39**(4):1294-1309.

100. Reichenbach B, Göpel Y, Görke B: **Dual control by perfectly overlapping sigma 54- and sigma 70- promoters adjusts small RNA GlmY expression to different environmental signals.** *Mol Microbiol* 2009, **74**(5):1054-1070.

101. Kalamorz F, Reichenbach B, März W, Rak B, Görke B: **Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in** *Escherichia coli*. *Mol Microbiol* 2007, **65**(6):1518-1533.

102. Sperandeo P, Lau FK, Carpentieri A, De Castro C, Molinaro A, Dehò G, Silhavy TJ, Polissi A: **Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of** *Escherichia coli*. *J Bacteriol* 2008, **190**(13):4460-4469.

103. Tokuda H: **Biogenesis of outer membranes in Gram-negative bacteria.** *Biosci Biotechnol Biochem* 2009, **73**(3):465-473.

104. Tran AX, Dong C, Whitfield C: **Structure and functional analysis of LptC, a conserved membrane protein involved in the lipopolysaccharide export pathway in** *Escherichia coli*. *J Biol Chem* 2010, **285**(43):33529-33539.

105. Sperandeo P, Pozzi C, Deho G, Polissi A: **Non-essential KDO biosynthesis and new essential cell envelope biogenesis genes in the** *Escherichia coli* **yrbG-yhbG locus.** *Res Microbiol* 2006, **157**(6):547-558.

106. Ueta M, Ohniwa RL, Yoshida H, Maki Y, Wada C, Wada A: **Role of HPF (hibernation promoting factor) in translational activity in** *Escherichia coli*. *J Biochem* 2008, **143**(3):425-433.

107. Zorrilla S, Chaix D, Ortega A, Alfonso C, Doan T, Margeat E, Rivas G, Aymerich S, Declerck N, Royer CA: **Fructose-1,6-bisphosphate acts both as an inducer and as a structural cofactor of the central glycolytic genes repressor (CggR).** *Biochemistry* 2007, **46**(51):14996-15008.

108. Letunic I, Yamada T, Kanehisa M, Bork P: **iPath: interactive exploration of biochemical pathways and networks.** *Trends Biochem Sci* 2008, **33**(3):101-103.

109. Kumar AS, Mody K, Jha B: **Bacterial exopolysaccharides - a perception.** *J Basic Microbiol* 2007, **47**(2):103-117.

110. Kovacs-Simon A, Titball RW, Michell SL: **Lipoproteins of bacterial pathogens.** *Infect Immun* 2011, **79**(2):548-561.

111. Swoboda JG, Campbell J, Meredith TC, Walker S: **Wall teichoic acid function, biosynthesis, and inhibition.** *Chembiochem* 2010, **11**(1):35-45.

112. Raetz CR, Reynolds CM, Trent MS, Bishop RE: **Lipid A modification systems in gram-negative bacteria.** *Annu Rev Biochem* 2007, **76**:295-329.

113. Barreteau H, Kovač A, Boniface A, Sova M, Gobec S, Blanot D: **Cytoplasmic steps of peptidoglycan biosynthesis.** *FEMS Microbiol Rev* 2008, **32**(2):168-207.

114. Jann K, Jann B: **Assembly of cellular surface structures.** In *Biology of the prokaryotes.* Edited by: Lengeler JW, Drews G, Schlegel HG. Stuttgart: Thieme Verlag; 1999:555-569.

115. Jacobi S, Schade R, Heuner K: **Characterization of the alternative sigma factor sigma54 and the transcriptional regulator FleQ of** *Legionella pneumophila*, **which are both involved in the regulation cascade of flagellar gene expression.** *J Bacteriol* 2004, **186**(9):2540-2547.

116. Kamal N, Dorrell N, Jagannathan A, Turner SM, Constantinidou C, Studholme DJ, Marsden G, Hinds J, Laing KG, Wren BW, *et al*: **Deletion of a previously uncharacterized flagellar-hook-length control gene** *fliK* **modulates the sigma54-dependent regulon in** *Campylobacter jejuni*. *Microbiology* 2007, **153**(Pt 9):3099-3111.

117. Poggio S, Osorio A, Dreyfus G, Camarena L: **The flagellar hierarchy of** *Rhodobacter sphaeroides* **is controlled by the concerted action of two enhancer-binding proteins.** *Mol Microbiol* 2005, **58**(4):969-983.

118. Raberg M, Reinecke F, Reichelt R, Malkus U, König S, Pötter M, Fricke WF, Pohlmann A, Voigt B, Hecker M, *et al*: *Ralstonia eutropha* **H16 flagellation changes according to nutrient supply and state of poly(3-hydroxybutyrate) accumulation.** *Appl Environ Microbiol* 2008, **74**(14):4477-4490.

119. Anderson JK, Smith TG, Hoover TR: **Sense and sensibility: flagellum-mediated gene regulation.** *Trends Microbiol* 2010, **18**(1):30-37.

120. Cavalier-Smith T: **Rooting the tree of life by transition analyses.** *Biol Direct* 2006, **1**:19.

121. Cavalier-Smith T: **Deep phylogeny, ancestral groups and the four ages of life.** *Philos Trans R Soc Lond B Biol Sci* 2010, **365**(1537):111-132.

122. Keleti G, Sykora JL: **Production and properties of cyanobacterial endotoxins.** *Appl Environ Microbiol* 1982, **43**(1):104-109.

123. Cullen TW, Trent MS: **A link between the assembly of flagella and lipooligosaccharide of the Gram-negative bacterium** *Campylobacter jejuni*. *Proc Natl Acad Sci USA* 2010, **107**(11):5160-5165.

124. Kyriakidis DA, Tiligada E: **Signal transduction and adaptive regulation through bacterial two-component systems: the** *Escherichia coli* **AtoSC paradigm.** *Amino Acids* 2009, **37**(3):443-458.

125. Comas I, González-Candelas F, Zúñiga M: **Unraveling the evolutionary history of the phosphoryl-transfer chain of the phosphoenolpyruvate: phosphotransferase system through phylogenetic analyses and genome context.** *BMC Evol Biol* 2008, **8**:147.

126. Barabote RD, Saier MH Jr: **Comparative genomic analyses of the bacterial phosphotransferase system.** *Microbiol Mol Biol Rev* 2005, **69**(4):608-634.

127. Mao XJ, Huo YX, Buck M, Kolb A, Wang YP: **Interplay between CRP-cAMP and PII-Ntr systems forms novel regulatory network between carbon metabolism and nitrogen assimilation in Escherichia coli.** *Nucleic Acids Res* 2007, **35**(5):1432-1440.

128. Houot L, Chang S, Pickering BS, Absalon C, Watnick PI: **The phosphoenolpyruvate phosphotransferase system regulates** *Vibrio cholerae* **biofilm formation through multiple independent pathways.** *J Bacteriol* 2010, **192**(12):3055-3067.

129. Lazazzera BA: **The phosphoenolpyruvate phosphotransferase system: as important for biofilm formation by** *Vibrio cholerae* **as it is for metabolism in** *Escherichia coli*. *J Bacteriol* 2010, **192**(16):4083-4085.

130. Martin-Verstraete I, Charrier V, Stülke J, Galinier A, Erni B, Rapoport G, Deutscher J: **Antagonistic effects of dual PTS-catalysed phosphorylation on the** *Bacillus subtilis* **transcriptional activator LevR.** *Mol Microbiol* 1998, **28**(2):293-303.

131. Lai X, Ingram LO: **Cloning and sequencing of a cellobiose phosphotransferase system operon from** *Bacillus stearothermophilus* **XL-65-6 and functional expression in** *Escherichia coli*. *J Bacteriol* 1993, **175**(20):6441-6450.

132. Yebra MJ, Viana R, Monedero V, Deutscher J, Pérez-Martinez G: **An esterase gene from** *Lactobacillus casei* **cotranscribed with genes encoding a phosphoenolpyruvate:sugar phosphotransferase system and regulated by a LevR-like activator and sigma54 factor.** *J Mol Microbiol Biotechnol* 2004, **8**(2):117-128.

133. Milenbachs Lukowiak A, Mueller KJ, Freitag NE, Youngman P: **Deregulation of** *Listeria monocytogenes* **virulence gene expression by two distinct and semi-independent pathways.** *Microbiology* 2004, **150**(Pt 2):321-333.

134. Plumbridge J: **An alternative route for recycling of N-acetylglucosamine from peptidoglycan involves the N-acetylglucosamine phosphotransferase system in** *Escherichia coli*. *J Bacteriol* 2009, **191**(18):5641-5647.

135. Born TL, Blanchard JS: **Structure/function studies on enzymes in the diaminopimelate pathway of bacterial cell wall biosynthesis.** *Curr Opin Chem Biol* 1999, **3**(5):607-613.

136. Robinson JB, Bauer WD: **Relationships between C4 dicarboxylic acid transport and chemotaxis in** *Rhizobium meliloti*. *J Bacteriol* 1993, **175**(8):2284-2291.

137. Yurgel SN, Kahn ML: **Dicarboxylate transport by rhizobia.** *FEMS Microbiol Rev* 2004, **28**(4):489-501.

138. Belitsky BR, Sonenshein AL: **An enhancer element located downstream of the major glutamate dehydrogenase gene of** *Bacillus subtilis*. *Proc Natl Acad Sci USA* 1999, **96**(18):10290-10295.

139. Dong T, Yu R, Schellhorn H: **Antagonistic regulation of motility and transcriptome expression by RpoN and RpoS in** *Escherichia coli*. *Mol Microbiol* 2011, **79**(2):375-386.

140. Heermann R, Jung K: **The complexity of the 'simple' two-component system KdpD/KdpE in** *Escherichia coli*. *FEMS Microbiol Lett* 2010, **304**(2):97-106.

141. Kearns DB: **A field guide to bacterial swarming motility.** *Nat Rev Microbiol* 2010, **8**(9):634-644.

142. Karatan E, Watnick P: **Signals, regulatory networks, and materials that build and break bacterial biofilms.** *Microbiol Mol Biol Rev* 2009, **73(2)**:310-347.

143. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, *et al*: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2010, , **38 Database**: D5-16.

144. UniprotConsortium: **The Universal Protein Resource (UniProt) in 2010.** *Nucleic Acids Res* 2010, , **38 Database**: D142-148.

145. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M: **From genomics to chemical genomics: new developments in KEGG.** *Nucleic Acids Res* 2006, , **34 Database**: D354-357.

146. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25(17)**:3389-3402.

147. Eddy SR: **A new generation of homology search tools based on probabilistic inference.** *Genome Inform* 2009, **23(1)**:205-211.

148. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, *et al*: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23(21)**:2947-2948.

149. van der Heijden RT, Snel B, van Noort V, Huynen MA: **Orthology prediction at scalable resolution by phylogenetic tree analysis.** *BMC Bioinformatics* 2007, **8**:83.

150. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R: **Dendroscope: An interactive viewer for large phylogenetic trees.** *BMC Bioinformatics* 2007, **8**:460.

151. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14(6)**:1188-1190.

152. Viégas FB, Wattenberg M, Feinberg J: **Participatory visualization with Wordle.** *IEEE Trans Vis Comput Graph* 2009, **15(6)**:1137-1144.

153. Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ, *et al*: **Assessing computational tools for the discovery of transcription factor binding sites.** *Nat Biotechnol* 2005, **23(1)**:137-144.

154. van Nimwegen E: **Finding regulatory elements and regulatory motifs: a general probabilistic framework.** *BMC Bioinformatics* 2007, **8(Suppl 6)**:S4.

155. Gershenzon NI, Stormo GD, Ioshikhes IP: **Computational technique for improvement of the position-weight matrices for the DNA/protein binding sites.** *Nucleic Acids Res* 2005, **33(7)**:2290-2301.

156. Mrázek J: **Finding sequence motifs in prokaryotic genomes–a brief practical guide for a microbiologist.** *Brief Bioinform* 2009, **10(5)**:525-536.

157. Claverie JM, Audic S: **The statistical significance of nucleotide position-weight matrix matches.** *Comput Appl Biosci* 1996, **12(5)**:431-439.

158. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS: **MEME SUITE: tools for motif discovery and searching.** *Nucleic Acids Res* 2009, , **37 Web Server**: W202-208.

159. Miwa Y, Nakata A, Ogiwara A, Yamamoto M, Fujita Y: **Evaluation and characterization of catabolite-responsive elements (cre) of *Bacillus subtilis*.** *Nucleic Acids Res* 2000, **28(5)**:1206-1210.

160. Molle V, Fujita M, Jensen ST, Eichenberger P, González-Pastor JE, Liu JS, Losick R: **The Spo0A regulon of *Bacillus subtilis*.** *Mol Microbiol* 2003, **50(5)**:1683-1701.

161. Hauser F, Pessi G, Friberg M, Weber C, Rusca N, Lindemann A, Fischer HM, Hennecke H: **Dissection of the *Bradyrhizobium japonicum* NifA+sigma54 regulon, and identification of a ferredoxin gene (*fdxN*) for symbiotic nitrogen fixation.** *Mol Genet Genomics* 2007, **278(3)**:255-271.

162. Kunisawa T: **Dichotomy of major bacterial phyla inferred from gene arrangement comparisons.** *J Theor Biol* 2006, **239(3)**:367-375.

163. Herlemann DP, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, Hugenholtz P, Brune A: **Genomic analysis of "*Elusimicrobium minutum*," the first cultivated representative of the phylum "*Elusimicrobia*" (formerly termite group 1).** *Appl Environ Microbiol* 2009, **75(9)**:2841-2849.

164. Gupta RS, Griffiths E: **Critical issues in bacterial phylogeny.** *Theor Popul Biol* 2002, **61(4)**:423-434.

165. Zhang H, Sekiguchi Y, Hanada S, Hugenholtz P, Kim H, Kamagata Y, Nakamura K: ***Gemmatimonas aurantiaca* gen. nov., sp. nov., a gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum *Gemmatimonadetes* phyl. nov.** *Int J Syst Evol Microbiol* 2003, **53(Pt 4)**:1155-1163.

166. Kunisawa T: **Evaluation of the phylogenetic position of the sulfate-reducing bacterium *Thermodesulfovibrio yellowstonii* (phylum *Nitrospirae*)** by means of gene order data from completely sequenced genomes. *Int J Syst Evol Microbiol* 2010, **60(Pt 5)**:1090-1102.

167. Zhaxybayeva O, Swithers KS, Lapierre P, Fournier GP, Bickhart DM, DeBoy RT, Nelson KE, Nesbø CL, Doolittle WF, Gogarten JP, *et al*: **On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales.** *Proc Natl Acad Sci USA* 2009, **106(14)**:5865-5870.

168. Sutcliffe IC: **Cell envelope architecture in the *Chloroflexi*: a shifting frontline in a phylogenetic turf war.** *Environ Microbiol* 2011, **13(2)**:279-282.

169. Botero LM, Brown KB, Brumefield S, Burr M, Castenholz RW, Young M, McDermott TR: ***Thermobaculum terrenum* gen. nov., sp. nov.: a non-phototrophic gram-positive thermophile representing an environmental clone group related to the Chloroflexi (green non-sulfur bacteria) and Thermomicrobia.** *Arch Microbiol* 2004, **181(4)**:269-277.

170. Thompson BG, Murray RG: **Isolation and characterization of the plasma membrane and the outer membrane of *Deinococcus radiodurans* strain Sark.** *Can J Microbiol* 1981, **27(7)**:729-734.

171. Saiki J, Kobayashi Y, Kawagoe K, Beppu T: ***Dictyoglomus thermophilum* gen. nov., sp. nov., a chemoorganotrophic, anaerobic, thermophilic bacterium.** *Int J Syst Bacteriol* 1985, **35(3)**:253-259.

172. Liebl W, Winterhalter C, Baumeister W, Armbrecht M, Valdez M: **Xylanase attachment to the cell wall of the hyperthermophilic bacterium *Thermotoga maritima*.** *J Bacteriol* 2008, **190(4)**:1350-1358.

173. Taylor M, Butler R, Chambers S, Casimiro M, Badii F, Merrick M: **The RpoN-box motif of the RNA polymerase sigma factor sigma N plays a role in promoter recognition.** *Mol Microbiol* 1996, **22(5)**:1045-1054.

174. Doucleff M, Pelton JG, Lee PS, Nixon BT, Wemmer DE: **Structural basis of DNA recognition by the alternative sigma-factor, sigma54.** *J Mol Biol* 2007, **369(4)**:1070-1078.

175. Burrows PC, Severinov K, Ishihama A, Buck M, Wigneshweraraj SR: **Mapping sigma 54-RNA polymerase interactions at the -24 consensus promoter element.** *J Biol Chem* 2003, **278(32)**:29728-29743.

176. Hutcheson SW, Bretz J, Sussan T, Jin S, Pak K: **Enhancer-binding proteins HrpR and HrpS interact to regulate hrp-encoded type III protein secretion in *Pseudomonas syringae* strains.** *J Bacteriol* 2001, **183(19)**:5589-5598.

177. Kim YK, McCarter LL: **Cross-regulation in *Vibrio parahaemolyticus*: compensatory activation of polar flagellar genes by the lateral flagellar regulator LafK.** *J Bacteriol* 2004, **186(12)**:4014-4018.

178. Koo IC, Stephens RS: **A developmentally regulated two-component signal transduction system in *Chlamydia*.** *J Biol Chem* 2003, **278(19)**:17314-17319.

179. Brahmachary P, Dashti MG, Olson JW, Hoover TR: ***Helicobacter pylori* FlgR is an enhancer-independent activator of sigma54-RNA polymerase holoenzyme.** *J Bacteriol* 2004, **186(14)**:4535-4542.

180. Cho K, Treuner-Lange A, O'Connor KA, Zusman DR: **Developmental aggregation of *Myxococcus xanthus* requires *frgA*, an *frz*-related gene.** *J Bacteriol* 2000, **182(23)**:6614-6621.

181. Ouyang Z, Blevins JS, Norgard MV: **Transcriptional interplay among the regulators Rrp2, RpoN and RpoS in *Borrelia burgdorferi*.** *Microbiology* 2008, **154(Pt 9)**:2641-2658.

182. Méndez-Ortiz MM, Hyodo M, Hayakawa Y, Membrillo-Hernández J: **Genome-wide transcriptional profile of *Escherichia coli* in response to high levels of the second messenger 3′,5′-cyclic diguanylic acid.** *J Biol Chem* 2006, **281(12)**:8090-8099.

183. Kim HS, Park SJ, Lee KH: **Role of NtrC-regulated exopolysaccharides in the biofilm formation and pathogenic interaction of *Vibrio vulnificus*.** *Mol Microbiol* 2009, **74(2)**:436-453.

184. Petrova OE, Sauer K: **A novel signaling network essential for regulating *Pseudomonas aeruginosa* biofilm development.** *PLoS Pathog* 2009, **5(11)**: e1000668.

185. Reading NC, Rasko DA, Torres AG, Sperandio V: **The two-component system QseEF and the membrane protein QseG link adrenergic and stress sensing to bacterial pathogenesis.** *Proc Natl Acad Sci USA* 2009, **106(14)**:5889-5894.

186. Ritchings BW, Almira EC, Lory S, Ramphal R: **Cloning and phenotypic characterization of *fleS* and *fleR*, new response regulators of *Pseudomonas aeruginosa* which regulate motility and adhesion to mucin.** *Infect Immun* 1995, **63(12)**:4868-4876.

187. Silbersack J, Jürgen B, Hecker M, Schneidinger B, Schmuck R, Schweder T: **An acetoin-regulated expression system of *Bacillus subtilis*.** *Appl Microbiol Biotechnol* 2006, **73(4)**:895-903.

188. Tucker NP, Ghosh T, Bush M, Zhang X, Dixon R: **Essential roles of three enhancer sites in sigma54-dependent transcription by the nitric oxide sensing regulatory protein NorR.** *Nucleic Acids Res* 2010, **38(4)**:1182-1194.

189. Palacios S, Starai VJ, Escalante-Semerena JC: **Propionyl coenzyme A is a common intermediate in the 1,2-propanediol and propionate catabolic pathways needed for expression of the** *prpBCDE* **operon during growth of** *Salmonella enterica* **on 1,2-propanediol.** *J Bacteriol* 2003, **185(9)**:2802-2810.

190. Yoshida A, Nishimura T, Kawaguchi H, Inui M, Yukawa H: **Enhanced hydrogen production from formic acid by formate hydrogen lyase-overexpressing** *Escherichia coli* **strains.** *Appl Environ Microbiol* 2005, **71(11)**:6762-6768.

191. Nakajima H, Takatani N, Yoshimitsu K, Itoh M, Aono S, Takahashi Y, Watanabe Y: **The role of the Fe-S cluster in the sensory domain of nitrogenase transcriptional activator VnfA from** *Azotobacter vinelandii.* *FEBS J* 2010, **277(3)**:817-832.